Generating Fuzzy Semantic Metadata Describing Spatial Relations from Images using the R-Histogram

Yuhang Wang, Fillia Makedon, James Ford, Li Shen
Department of Computer Science
Dartmouth College
Hanover, NH 03755
{wyh,makedon,jford,li}@cs.dartmouth.edu

Dina Goldin
Computer Science & Engineering Dept.
University of Connecticut
Storrs, CT 06269
dqg@cse.uconn.edu

ABSTRACT

Automatic generation of semantic metadata describing spatial relations is highly desirable for image digital libraries. Relative spatial relations between objects in an image convey important information about the image. Because the perception of spatial relations is subjective, we propose a novel framework for automatic metadata generation based on fuzzy k-NN classification that generates fuzzy semantic metadata describing spatial relations between objects in an image. For each pair of objects of interest, the corresponding R-histogram is computed and used as input for a set of fuzzy k-NN classifiers. The R-histogram is a quantitative representation of spatial relations between two objects. The outputs of the classifiers are soft class labels for each of the following eight spatial relations: 1) LEFT OF, 2) RIGHT OF, 3) ABOVE, 4) BELOW, 5) NEAR, 6) FAR, 7) INSIDE, 8) OUTSIDE. Because the classifier-training stage involves annotating the training images manually, it is desirable to use as few training images as possible. To address this issue, we applied existing prototype selection techniques and also devised two new extensions. We evaluated the performance of different fuzzy k-NN algorithms and prototype selection algorithms empirically on both synthetic and real images. Preliminary experimental results show that our system is able to obtain good annotation accuracy (92%–98% on synthetic images and 82%–93% on real images) using only a small training set (4–5 images).

Categories and Subject Descriptors

H.3.7 [**Information Storage And Retrieval**]: Digital Libraries; H.2.8 [**Database Management**]: Database Applications—*Image databases*

General Terms

Design, Management

Keywords

Metadata, Spatial Relations, Image Digital Library, k-Nearest Neighbor Rule, Prototype Selection, R-Histogram

1. INTRODUCTION

The emergence of digital libraries makes image retrieval an issue of rapidly increasing importance. There are two major approaches: text metadata-based image retrieval and content-based image retrieval. The first of these requires that images be annotated with text metadata, while the second applies image processing and pattern recognition techniques to extract visual features (color, texture, shape, and spatial relations) from images [10, 28, 34]. Similarity of images is then determined using distance functions defined on these features. These two approaches are complimentary: text metadata-based approaches can support keyword-based queries, and content-based approaches can support query-by-example. The text metadata for images can be created by manual annotation; however, manual annotation is expensive and time-consuming, and thus infeasible for large-scale image digital libraries. Clearly, automatic metadata generation systems are desirable. In this paper, we focus on the automatic generation of fuzzy semantic metadata describing spatial relations from segmented images. We will explain the meaning of "fuzzy semantic metadata describing spatial relations" in the following section. We assume the images are segmented and each object is assigned a unique label; i.e., we deal with symbolic images, as defined formally in [13].

1.1 Related Work in Spatial Relation Representations

The determination of spatial relations between objects in an image is an important component of image content description and access. For example, the spatial relationship between brain lesions and anatomical brain structures in medical images is critically important for early disease diagnosis and thus important for image retrieval. Researchers in different fields have worked on the analysis of spatial relations among and between objects. Many approaches have been proposed for qualitative spatial relations [7] and fuzzy spatial relations [4]. A widely adopted idea for modeling spatial relations was proposed by Freeman [11]. In this scheme, the primitive spatial relations between two objects are: 1) LEFT OF, 2) RIGHT OF, 3) ABOVE, 4) BELOW, 5) BEHIND, 6) IN FRONT OF, 7) BESIDE, 8) NEAR, 9)

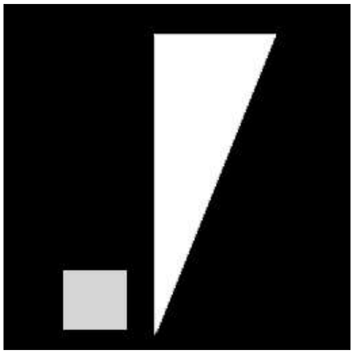


Figure 1: Example image illustrating the subjectiveness of human perception of spatial relations.

FAR, 10) TOUCHING, 11) BETWEEN, 12) INSIDE, and 13) OUTSIDE. The first six are called the primitive directional relations. Freeman also proposed that a fuzzy degree of truth be associated with each spatial relation. Each spatial relation thus defined gives a distinct semantic meaning. Therefore, we consider the fuzzy degree of truth with its associated spatial relation as fuzzy semantic metadata describing spatial relations, or in short, *fuzzy spatial relation metadata* (*FSR-metadata*).

In the past, FSR-metadata were automatically generated by:

1. Specially-designed algorithms. These methods include the compatibility method [27, 26], the angle aggregation method [19], the effective force method [23] and the morphological approach [3]. All of these methods are not based on machine learning and do not pay attention to properties of human perceptions. In fact, these methods have shown to often give inconsistent values of fuzzy spatial relations for the same images [35].

2. Learning-based methods. Surprisingly, there is less reported work for this category of approaches in relation to the problem in hand. Wang and Keller [35] proposed a neural network based system to generate FSR-metadata for images. Fuzzy directional relationship metadata values were generated from neural networks fed by angle histograms. These networks were trained on aggregate responses from a panel of people. However, in this work, the training images were chosen arbitrarily and the system only generated FSR-metadata related to the four primitive directional relationships: 1) LEFT OF, 2) RIGHT OF, 3) ABOVE, 4) BELOW.

All the above approaches derive a single "standard" set of FSR-metadata. However, since the perception of spatial relations is highly subjective, it can vary significantly from person to person. For example, one person may assign degree of truth 0.9 to the FSR-metadata "the triangle is above the rectangle" in the image shown in Figure 1, meanwhile another person may think differently and assign 0.6. When a user queries a digital library containing only one single set of "standard" FSR-metadata, she has to learn the mapping from her own idea of spatial relations to the standard metadata values. We believe that the system should learn from a user's peculiarity but not vice versa. Therefore, it is desirable to derive *personalized FSR-metadata* that are consistent with user's intuitive values. Coincidentally, the need for personalized search was also emphasized by McKeown *et al.* in [25] from another perspective.

Many families of fuzzy directional relations rely on the construction of the histogram of angles [27, 26, 19, 35] or the Force Histogram [24, 23]. Given an object of interest A and a reference object R, the histogram of angles is computed from the angle between any two points in both objects and normalized by the maximum frequency. The Force Histogram (F-histogram) generalizes the histogram of angles. It is computed from the sums of the elementary forces that are exerted by the points of A on those of R in each direction. However, all these angle histogram approaches can only represent directional relations, but not the topological spatial relations "inside" and "outside" [36]. Moreover, distance information is not explicitly taken into account and it is impossible to extract metadata for the FAR and NEAR spatial relations of the Freeman model. In many cases, quantitative representation of spatial relations taking into account shape, size, orientation and distance is required. To address these issues, the R-histogram was proposed by Wang and Makedon in [36]. This earlier work constitutes one of the building blocks of the metadata generation system discussed in this paper.

1.2 Related Work in Automatic Metadata Generation

Many methods have been proposed for automatic metadata generation [30, 12, 32, 22]. Because machine learning based methods are robust and adaptable, these methods are gaining popularity, *e.g.* Support Vector Machines [14], Hidden Markov Models [33], and statistical methods [16]. Many authors have suggested that automation of metadata generation can be addressed as a classification problem [6, 14]. In the case of an image digital library, Wang and Li [33] reported on the development of a content-based image retrieval system that "trains" computers to assign keywords to images based on automatically created dictionaries of concepts.

One problem with this classification approach is that the labels for the training data set must be generated manually, and this process is rather expensive. Therefore, in an automatic metadata generation system based on classification, it is desirable that the size of the training data set be small. Clearly, systematic approaches for reducing the size of the training data set by choosing the most representative training data would be useful.

1.3 Our Approach

In summary, the following main questions have not been addressed by previous work in automatic generation of FSR-metadata:

1. Generating metadata describing not only directional relations, but also the spatial relations NEAR and FAR, plus the topological spatial relations INSIDE and OUTSIDE.

2. Incorporating methods to systematically reduce the size of the training image set.

3. Providing both personalized and "standard" metadata for spatial relations.

In this paper, we address the above questions by describing a learning-based system that extracts FSR-metadata by training a set of fuzzy k nearest neighbor (k-NN) classifiers. Using R-histograms as the input of the k-NN classifiers, the system generates metadata for both directional relations LEFT OF, RIGHT OF, ABOVE, BELOW and the spatial relations NEAR, FAR, INSIDE, OUTSIDE. To reduce the number of training images, we apply two types of prototype selection methods: editing and condensing [20]. The goal is to select the most "important" training images and use them as the prototypes in the k-NN classifiers. The system is able to provide both a) "standard" metadata for spatial relations (by aggregating the responses of all users or groups of users) and b) personalized metadata (obtained through training by each individual user).

1.4 Outline of The Paper

The remainder of the paper is organized as follows: Section 2 gives an overview of the proposed system; Section 3 presents the details of our fuzzy spatial relation metadata generation methods; Section 4 demonstrates the preliminary experimental results of our system on synthetic images and real images; Section 5 concludes the paper.

2. SYSTEM OVERVIEW

An overview of PerGen, our personalized metadata generation system based on fuzzy k-NN classification, is illustrated in Figure 2. Our approach is based on R-histograms computed for each pair of objects of interest in input images. We use a fuzzy k-NN classifier for each of the four spatial relation pairs "LEFT OF/RIGHT OF," "ABOVE/BELOW," "NEAR/FAR," and "INSIDE/OUTSIDE." The first two classifiers operate directly on the R-histograms. For the remaining two classifiers, we extract higher level feature vectors from R-histograms and feed them to the classifiers to reduce dimensionality and incorporate prior knowledge about R-Histograms. In training these fuzzy k-NN classifiers, we consider the properties of human perception in the form of soft class labels. These are obtained from user responses to a small set of representative images, which are the result of prototype selection (explained below in Section 3.4). The outputs of the classifiers are the fuzzy values associated with corresponding FSR-metadata. The soft labels assigned by a user are stored in her user profile to facilitate *personalized metadata generation.* Soft labels from all users can also be combined for generation of "standard" FSR-metadata.

3. FUZZY SEMANTIC METADATA GENERATION

3.1 The R-Histogram

Given a reference object R and an object of interest A, the goal is to represent, quantitatively, the spatial relations between R and A. Consider the vector originating from a pixel x on the boundary of R to a pixel y on the boundary of A. If x and y don't coincide, the angle between the x-axis of the coordinate frame and $\overrightarrow{xy}$ is computed. This angle, denoted by $\theta(x, y)$, takes values in $[-\pi, \pi]$. The labeled distance from x to y, denoted by $LD(x, y)$, is defined as a

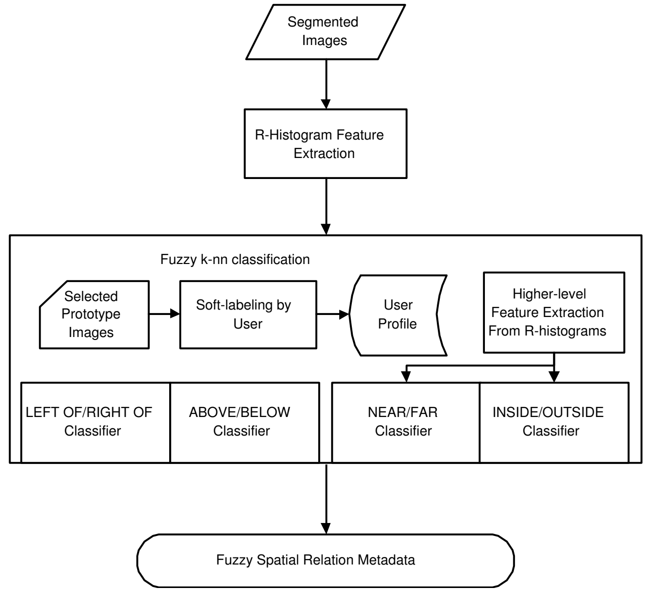


Figure 2: Personalized FSR-metadata generation system based on classification.

Table 1: Labels in the Labeled Distance

Pixel x inside A	Pixel y inside R	$l(x, y)$
False	False	0
False	True	1
True	False	2
True	True	3

pair $(d(x, y), l(x, y))$, where $d(x, y)$ is the Euclidean distance from x to y and $l(x, y)$ is defined in Table 1.

For the set of vectors originating from any pixel on the boundary of R to any pixel on the boundary of A, a histogram is constructed as follows: Let x and y be pixels on the boundaries of R and A respectively. The bin $H(I, J, L)$ is incremented as follows:

$$H(I,J,L) = \begin{cases} H(I,J,L)+1 & \text{if } \theta(x,y) \in A_I \\ & \quad \wedge\ (d(x,y) \in D_J \\ & \quad \wedge\ l(x,y) = L \\ H(I,J,L) & \text{otherwise} \end{cases} \quad (1)$$

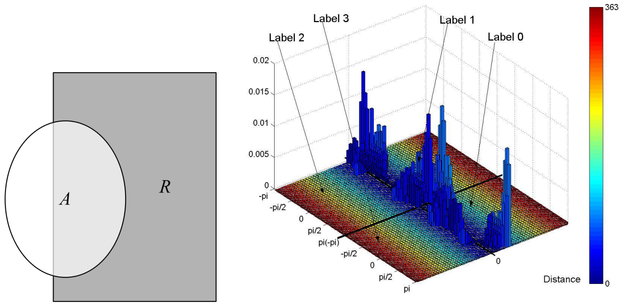
where A_I is the range of angle values spanned by bin $H(I, J, L)$, D_J is the range of distance values spanned by bin $H(I, J, L)$, and $L \in \{0, 1, 2, 3\}$ is the label associated with the distance values spanned by bin $H(I, J, L)$.

Then the histogram is normalized as follows:

$$h(I,J,L) = \frac{H(I,J,L)}{\sum_{I'=1}^{n_A} \sum_{J'=1}^{n_D} \sum_{L'=0}^{3} H(I',J',L')} \quad (2)$$

where n_A is the number of angle bins and n_D the number of distance bins. The normalized histogram, denoted as $RH(A, R)$, is defined to be the R-Histogram of object A relative to object R.

An example of the R-Histogram is illustrated in Figure 3(b), where the x-axis is associated with angles and the y-axis with distances. Each of the four quadrants shows a portion of the histogram with one of the four labels asso-



(a) (b)

Figure 3: (a) The gray rectangle R is the reference object and the white ellipse A is the object of interest. Note that the two object overlap. (b) The R-histogram for the two objects. Each quadrant is associated with a unique label.

ciated with the distances. $RH(A, R)$ is invariant to translation. When a rotation is applied to A and R, $RH(A, R)$ is simply shifted along the x-axis. $RH(A, R)$ can also be scale-invariant if the distances are normalized by dividing them by the maximum distance. $RH(A, R)$ encapsulates rich information about the spatial relations between A and R.

To model the spatial relations of *multiple objects* in an image, we can use R-Histograms as the arc attributes in Attributed Relational Graphs (ARG).

3.2 Fuzzy k-NN Classification

In the proposed system, PerGen, we address the task of automatic generation of FSR-metadata as a fuzzy classification problem. There are many different types of fuzzy classifiers [20]. We chose to use the fuzzy k-NN classifier because it supports non-linear decision boundaries and prototype selection.

The k-NN classifier [8] is a well-known nonparametric classifier. Let X be a metric space with a global distance function (the metric d) which, for every two points x, y in X, gives the distance between them. Let $\Omega = \{\omega_1, \ldots, \omega_c\}$ be a set of class labels, where c is the number of classes. Let $V = \{(v_1, l(v_1)), \ldots, (v_n, l(v_n))\}$ be a labeled reference set, where $v_i \in X$ and $l(v_i) \in \Omega$. To classify a new input x, the k nearest prototypes are retrieved from V together with their class labels. The input x is labeled to the majority class label corresponding to the k nearest neighbors.

Fuzzy k-NN extends k-NN by replacing the crisp labels above with *soft labels*, $l(v_i) \in [0, 1]^c$. Different ways of combining the soft labels of the k nearest neighbors to compute the soft output label $l(x)$ for $x \in X$ have been proposed [17, 18, 8, 2]. In this work, we used the schemes proposed by Jóźwik [17] and Keller *et al.* [18]. To find the ith component of the soft class label $l_i(x)$ for x, Jozwik's method [17] simply averages the soft labels of the k nearest neighbors:

$$l_i(x) = \frac{1}{k} \sum_{j=1}^{k} l_i(z_j) \tag{3}$$

where
$z_1, \ldots, z_k$ are the k nearest neighbors of input point x.

The method by Keller *et al.* [18] takes distances into consideration:

$$l_i(x) = \frac{\sum_{j=1}^{k} l_i(z_j)(d_j)^{-\frac{2}{m-1}}}{\sum_{j=1}^{k}(d_j)^{-\frac{2}{m-1}}} \tag{4}$$

where
m is a "fuzzification" parameter;
d_j is the distance between x and its jth nearest neighbor z_j.

3.3 Distance Functions

One important component in k-NN classifier design is the choice of the distance function d. In [36], we used histogram intersection as the distance metric in a Query By Example (QBE) system. It is shown in [31] that when histograms are normalized, histogram intersection is given by

$$d(h_1, h_2) = \sum_{I=1}^{n_A} \sum_{J=1}^{n_D} \sum_{L=0}^{3} |h_1(I, J, L) - h_2(I, J, L)|$$

In this work, we show how to extend this distance function to create a variant d_1 for use with our k-NN classifiers for "LEFT OF/RIGHT OF" and "ABOVE/BELOW." The motivation behind this is to "separate" the directional component from the distance component in R-Histograms. Intuitively, d_1 is the minimum d obtained over all possible shifts of the second histogram along the distance axis (y-axis as shown in Figure 3(b)).

$$\begin{aligned} & d_1(h_1, h_2) \\ &= \min_{t \in [-n_D, n_D]} \sum_{I=1}^{n_A} \sum_{J=1}^{n_D} \sum_{L=0}^{3} |h_1(I, J, L) - h_2(I, J+t, L)| \end{aligned}$$

The distance function d_2 used by the classifier for "NEAR/FAR" is the difference between the weighted means over all distance bins of the two histograms in question.

$$a(h) = \sum_{J=1}^{n_D} J \cdot \sum_{I=1}^{n_A} \sum_{L=0}^{3} h(I, J, L)$$

$$d_2(h_1, h_2) = |a(h_1) - a(h_2)|$$

Here, $a(h)$ can be viewed as higher-level feature vectors extracted from h.

For the classifier for "INSIDE/OUTSIDE", we extract a higher-level feature vector $\vec{S}(h) = [s_1(h), s_2(h), s_3(h), s_4(h)]$ from the raw R-Histogram h to provide the input to the classifier, where $s_i(h)$ is the sum of the bin values of h in quadrant i. Then d_3 is defined as the L_1 distance between the extracted vectors.

$$d_3(h_1, h_2) = L_1(\vec{S}(h_1), \vec{S}(h_2))$$

This extraction step achieves two things: (a) it reduces the dimensionality of the feature space and (b) it incorporates prior knowledge about R-Histograms. For example, if A is completely inside R, then we know that the R-Histogram will only be populated in quadrant 2 and vice versa.

3.4 Reducing The Number of Training Images

To reduce the number of training images, we apply prototype selection methods to select the most representative training images. Only these are used as prototypes in the four PerGen k-NN classifiers.

Let $Z = \{(z_1, l(z_1)), \ldots, (z_n, l(z_n))\}$ be the whole labeled data set. Prototype selection consists of selecting a particular subset of prototypes and applying k-NN classification

Table 3: Number of prototypes $|V|$ and annotation accuracy using the fuzzy k-NN classifier by Keller *et al.* ($m = 2$), with different prototype selection methods

Methods			Left of/Right of		Above/Below		Near/Far		Inside/Outside	
distance function	k	prototype selection method	$\|V\|$	accuracy	$\|V\|$	accuracy	$\|V\|$	accuracy	$\|V\|$	accuracy
extended distance functions: d_1, d_2, d_3	1	No prototype selection	100	0.98	100	0.98	100	0.89	100	0.97
		YCFE	90	0.97	92	0.98	74	0.90	90	0.96
		FERS	5	0.97	5	0.98	5	0.91	5	0.97
		FC	10	0.98	6	0.97	17	0.90	36	0.97
		YCFE+FC	7	0.96	6	0.98	9	0.92	32	0.96
		YCFE+FERS	5	0.96	5	0.95	5	0.91	5	0.95
	2	No prototype selection	100	0.99	100	0.98	100	0.91	100	0.97
		YCFE	84	0.98	90	0.98	70	0.89	92	0.97
		FERS	5	0.99	5	0.97	5	0.9	5	0.96
		FC	13	0.95	6	0.78	18	0.91	36	0.97
		YCFE+FC	4	0.96	4	0.98	7	0.90	32	0.97
		YCFE+FERS	5	0.95	5	0.93	5	0.9	5	0.97
	3	No prototype selection	100	0.98	100	0.98	100	0.9	100	0.98
		YCFE	82	0.99	90	0.96	70	0.9	90	0.95
		FERS	5	0.88	5	0.86	5	0.92	5	0.97
		FC	14	0.95	12	0.93	17	0.9	35	0.95
		YCFE+FC	9	0.90	9	0.95	4	0.9	32	0.96
		YCFE+FERS	5	0.84	5	0.85	5	0.87	5	0.96
histogram intersection: d	1	No prototype selection	100	0.96	100	0.96	100	0.72	100	0.84
		YCFE	90	0.94	92	0.96	72	0.66	72	0.80
		FERS	5	0.96	5	0.96	5	0.60	5	0.80
		FC	10	0.96	8	0.86	22	0.68	17	0.84
		YCFE+FC	7	0.92	5	0.87	13	0.70	9	0.76
		YCFE+FERS	5	0.94	5	0.9	5	0.58	5	0.80
	2	No prototype selection	100	0.96	100	0.96	100	0.76	100	0.82
		YCFE	82	0.96	90	0.96	60	0.62	78	0.78
		FERS	5	0.98	5	0.92	5	0.55	5	0.80
		FC	13	0.86	10	0.88	23	0.74	15	0.80
		YCFE+FC	4	0.78	4	0.68	8	0.58	6	0.68
		YCFE+FERS	5	0.95	5	0.96	5	0.56	5	0.80
	3	No prototype selection	100	0.98	100	0.96	100	0.76	100	0.82
		YCFE	82	0.96	90	0.92	60	0.62	78	0.78
		FERS	5	0.70	5	0.60	5	0.52	5	0.80
		FC	16	0.94	12	0.86	26	0.78	18	0.88
		YCFE+FC	8	0.80	8	0.65	9	0.64	6	0.56
		YCFE+FERS	5	0.7	5	0.58	5	0.64	5	0.78

Table 2: Editing/condensing combinations

Methods	Description
YCFE	YCFE editing only
FERS	FERS editing only
FC	FC condensing only
YCFE+FC	YCFE editing, then FC condensing
YCFE+FERS	YCFE editing, then FERS editing

using only the samples selected. Two different strategies exist in the literature:

- Condensing algorithms find a minimal subset $V \subseteq Z$ that leads to approximately the same performance as the k-NN rule using the whole Z.
- Editing algorithms find a subset $V \subseteq Z$ by eliminating erroneously labeled prototypes from Z.

It has been shown that using editing algorithms in conjunction with condensing algorithms can produce a training set that is significantly smaller than the original with similar or better classification performance [9].

There are many editing and condensing algorithms for crisp k-NN classifiers [37, 15, 21] in the literature. However, much less work has been reported for fuzzy k-NN.

In this paper, we evaluated the following prototype selection methods and their combinations for fuzzy k-NN.

a) YCFE: The edited fuzzy k-NN classifier proposed by Yang and Chen [38]. This method is similar to Wilson's editing method defined for crisp k-NN [37]. The algorithm works as follows:

pick k_1 to be the number of neighbors for the training stage and the threshold α for the limit discrepancy between soft class labels.
for $q = 1$ to N **do**
    find the k_1 nearest neighbors of z_q using $Z - \{z_q\}$ as the reference set.
    calculate the estimated soft class label of z_q, $\hat{l}(z_q) = [\hat{l}_1(z_q), ..., \hat{l}_c(z_q)]$ using fuzzy k-NN method.

Table 4: Number of prototypes $|V|$ and annotation accuracy using the fuzzy k-NN classifier by Jóźwik, with different prototype selection methods

Methods			Left of/Right of		Above/Below		Near/Far		Inside/Outside	
distance function	k	prototype selection method	$\|V\|$	accuracy	$\|V\|$	accuracy	$\|V\|$	accuracy	$\|V\|$	accuracy
extended distance functions: d_1, d_2, d_3	1	No prototype selection	100	0.52	100	0.48	100	0.44		
		YCFE	40	0.48	34	0.44	54	0.44		
		FERS	5	0.72	5	0.78	5	0.44		
		FC	5	0.52	5	0.66	18	0.44		
		YCFE+FC	2	0.48	2	0.34	3	0.44		
		YCFE+FERS	5	0.48	5	0.46	5	0.44		
	2	No prototype selection	100	0.98	100	0.96	100	0.86	100	0.88
		YCFE	90	0.96	88	0.94	70	0.82	86	0.92
		FERS	5	0.90	5	0.98	5	0.86	56	0.88
		FC	7	0.80	10	0.98	17	0.86	6	0.80
		YCFE+FC	4	0.84	4	0.76	3	0.58	4	0.9
		YCFE+FERS	5	0.92	5	0.92	5	0.90	5	0.88
	3	No prototype selection	100	0.94	100	0.92	100	0.84	100	0.92
		YCFE	84	0.98	82	0.98	72	0.84	88	0.92
		FERS	5	0.52	5	0.51	5	0.78	5	0.86
		FC	12	0.56	13	0.76	13	0.86	8	0.88
		YCFE+FC	6	0.70	7	0.66	5	0.80	5	0.96
		YCFE+FERS	5	0.44	5	0.46	5	0.58	5	0.78
histogram intersection: d,	1	No prototype selection	100	0.52	100	0.48	100	0.44		
		YCFE	39	0.48	34	0.44	54	0.44		
		FERS	5	0.76	5	0.80	5	0.44		
		FC	6	0.82	8	0.64	20	0.44		
		YCFE+FC	2	0.48	3	0.88	3	0.44		
		YCFE+FERS	5	0.50	5	0.44	5	0.44		
	2	No prototype selection	100	0.96	100	0.96	100	0.64	100	0.8
		YCFE	90	0.96	88	0.96	66	0.52	76	0.8
		FERS	5	0.92	5	0.90	5	0.48	5	0.78
		FC	10	0.84	8	0.92	23	0.64	18	0.76
		YCFE+FC	5	0.74	5	0.67	11	0.38	4	0.8
		YCFE+FERS	5	0.92	5	0.94	5	0.46	5	0.78
	3	No prototype selection	100	0.92	100	0.94	100	0.68	100	0.86
		YCFE	84	0.98	82	0.98	56	0.52	80	0.8
		FERS	5	0.48	5	0.5	5	0.40	5	0.78
		FC	14	0.88	12	0.72	23	0.56	16	0.6
		YCFE+FC	6	0.70	7	0.52	6	0.20	10	0.7
		YCFE+FERS	5	0.46	5	0.40	5	0.38	5	0.78

    if $\left|\left|l(z_q) - \hat{l}(z_q)\right|\right| \geq \alpha$ **then**
        mark z_q for deletion.
    end if
end for
to obtain V, delete from Z all marked elements.
return V with the class labels.

b) FERS: We extended the algorithm for editing by random selection for crisp k-NN [29] to fuzzy k-NN.

pick the threshold α for the limit discrepancy between soft class labels for calculation of the error rate.
choose $v = |V|, c \leq v \leq N$. Pick the maximal number of steps T. Set $E = 1$.
for $i = 1$ to T **do**
    select at random $V \subset Z$, $|V| = v$.
    calculate the error rate $e(V)$ of the fuzzy k-NN classifier on $Z - \{V\}$, with V as the reference set.
    if $e(V) < E$ **then**
        store V and set $E = e(V)$.
    end if
end for
return V with the class labels.

c) FC: We extended the condensing algorithm in [1] for crisp k-NN to fuzzy k-NN.

pick the threshold α for the limit discrepancy between soft class labels.
pick the maximal number of steps T. Set $V = Z$.
for $i = 1$ to T **do**
    select at random $v \subset V$.
    find the k nearest neighbors of v using $V - \{v\}$ as the reference set.
    calculate the estimated soft class label of v, $\hat{l}(v) = [\hat{l}_1(v), ..., \hat{l}_c(v)]$ using fuzzy k-NN method.
    if $\left|\left|l(v) - \hat{l}(v)\right|\right| < \alpha$ **then**
        set $V = V - \{v\}$.
    end if
end for
return V with the class labels.

In the above algorithms, the norm metric $||\cdot||$ is the L_1 distance.

Table 5: Best methods for each of the four classifiers

Classifiers	Best Methods	$\|V\|$	accuracy
LEFT OF/RIGHT OF	YCFE+FC, Keller, m=2, k=2, using d_1 as distance function	4	0.96
	FERS, Keller, m=2, k=1, using d_1 as distance function	5	0.97
ABOVE/BELOW	YCFE+FC, Keller, m=2, k=2, using d_1 as distance function	4	0.98
	FERS, Keller, m=2, k=1, using d_1 as distance function	5	0.98
NEAR/FAR	FERS, Keller, m=2, k=3, using d_2 as distance function	5	0.92
	YCFE+FERS, Keller, m=2, k=1, using d_2 as distance function	5	0.91
	FERS, Keller, m=2, k=1, using d_2 as distance function	5	0.91
INSIDE/OUTSIDE	FERS, Keller, m=2, k=1, extended distance function	5	0.97
	YCFE+FERS, Keller, m=2, k=2, using d_3 as distance function	5	0.97
	FERS, Keller, m=2, k=1, using d_3 as distance function	5	0.97

Different combinations of the above algorithms are possible as shown in Table 2. The performance of these different combinations are evaluated experimentally in the next section.

4. EXPERIMENTAL RESULTS

We have implemented the above algorithms in MATLAB R13.1. To test the performance of different algorithms with different parameters and distance functions, experiments were conducted on synthetic images and real images.

4.1 Synthetic Images

The synthetic image data set used for our experiments consists of 200 images of size 256x256 containing rectangle and ellipse objects. The rectangles are always in the middle, while the locations of the ellipses are random, resulting in various relative spatial relations with the rectangle. In the tests, the rectangle is designated as the reference object and the ellipse the object of interest.

First, users assigned soft labels to each image by giving an appropriate value between 0 and 1 for each spatial relationship: 1) LEFT OF, 2) RIGHT OF, 3) ABOVE, 4) BELOW, 5) NEAR, 6) FAR, 7) INSIDE, 8) OUTSIDE. Users were instructed to give 0 to an image if the two objects do not satisfy the spatial relation at all, and to give 1 if the two objects satisfy the spatial relation perfectly, in the user's opinion. Note that each user was asked to annotate all the images because we need the labeled data set to evaluate the effectiveness of prototype selection. In a real-world implementation, once the prototypes are selected, a user would only have to annotate the small set of prototype images.

Ten unique experimental sets were created from the whole data set; each was created by randomly partitioning the 200 images into 100 training images and 100 test images. We consider a test image to be accurately annotated if the L_1 distance between the estimated soft label and the manually assigned label does not exceed a predefined threshold chosen by the user. This threshold was set to be 0.2 in the experiments. The threshold α used in YCFE and FERS were set to be 0.1, i.e. half of the user-defined error threshold.

Different combinations of prototype selection algorithms and different fuzzy k-NN classifiers using different distance functions and parameters were applied. For FERS, v was set at 5 and T was set to be 1000. For each trial, the number of selected prototype images and the classification accuracy were recorded. The results, averaged over ten trials, are shown in Table 3 and Table 4. Some cells of Table 4 are empty because the corresponding tests failed due to an empty prototype set.

We can observe from the results that:

- The fuzzy k-NN by Keller *et al.* outperformed Jóźwik's scheme;
- Good prototype selection algorithms can reduce the number of prototypes significantly (up to 96%) with no or small degradation in accuracy;
- The extended distance functions generally outperform simple histogram intersection;
- YCFE itself cannot reduce the number of prototypes significantly;
- No one single combinations works well for all classifiers. Table 5 summarizes the best methods for each of the four classifiers.

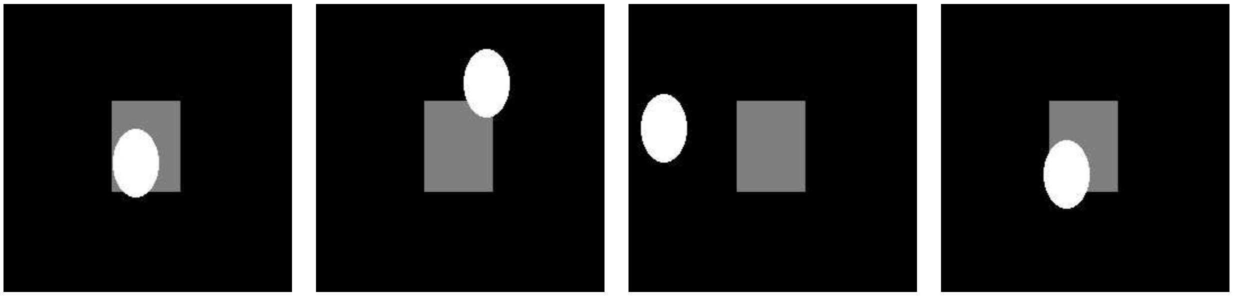
Four test images are shown in Figure 4. Their corresponding metadata assigned by the system and the user are listed in Table 6–9. Note that only the best results using the methods in Table 5 are shown. When the user-assigned metadata value is 0, the automatically-generated value is sometimes very close to, but not equal to, 0 (*e.g.* the row for "RIGHT OF" in Table 8). Although the difference is small, it is sometimes counter-intuitive. We will discuss possible ways to fix this problem in the next section.

4.2 Real Images

In order to evaluate the performance of the classifiers trained with synthetic images on real images, we tested with 10 photos of indoor and outdoor scenes. The objects in the images were segmented semi-manually with MATLAB. Due to space limitations, only two are shown here in Figure 5 and 6. Twenty-six object pairs from the 10 photos were investigated. The annotation accuracy rate obtained with the best classifiers is shown in Table 12.

5. CONCLUSION AND FUTURE WORK

The main contribution of this paper is to introduce a novel automatic metadata generation framework, PerGen, that uses fuzzy k-NN classification to generate fuzzy semantic metadata describing spatial relations for images. This framework is extensible and can be used to build digital libraries to allow searching and data mining of images based on spatial relations. A prototype system has been implemented and tested with both synthetic and real images. This paper also demonstrated the effectiveness of prototype selection algorithms in reducing the number of training images.



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

Figure 4: Example test images used in experiments

Table 6: FSR-metadata for Figure 4(a)

Spatial Relation	Values by	
	system	user
LEFT OF	0.47	0.5
RIGHT OF	0	0
ABOVE	0	0
BELOW	0.76	0.7
NEAR	1	1
FAR	1	0
INSIDE	0.92	0.9
OUTSIDE	0.08	0.1

Table 7: FSR-metadata for Figure 4(b)

Spatial Relation	Values by	
	system	user
LEFT OF	0	0
RIGHT OF	0.79	0.9
ABOVE	0.83	0.95
BELOW	0	0
NEAR	1	1
FAR	0	0
INSIDE	0.11	0.1
OUTSIDE	0.89	0.9

Table 8: FSR-metadata for Figure 4(c)

Spatial Relation	Values by	
	system	user
LEFT OF	0.81	1
RIGHT OF	0.03	0
ABOVE	0.68	0.6
BELOW	0	0
NEAR	0.06	0.1
FAR	0.94	0.9
INSIDE	0	0
OUTSIDE	1	1

Table 9: FSR-metadata for Figure 4(d)

Spatial Relation	Values by	
	system	user
LEFT OF	0.79	0.9
RIGHT OF	0	0
ABOVE	0	0
BELOW	1	0.8
NEAR	0.95	1
FAR	0.05	0
INSIDE	0.82	0.7
OUTSIDE	0.18	0.3

Figure 5: Test image showing a car on the roadside.

Table 10: FSR-metadata for Figure 5. The road is the reference object for the car. "Above" and "below" were not evaluated because they are not meaningful in this case.

Spatial Relation	Values by	
	system	user
LEFT OF	0.68	0.9
RIGHT OF	0.01	0
NEAR	1	1
FAR	0	0
INSIDE	0.41	0.5
OUTSIDE	0.59	0.5

Figure 6: Test image showing two vases.

Table 11: FSR-metadata for Figure 6. The taller vase is the reference object for the shorter one.

Spatial Relation	Values by	
	system	user
LEFT OF	0	0
RIGHT OF	1	1
ABOVE	0	0
BELOW	0.11	0.2
NEAR	0.37	0.5
FAR	0.63	0.5
INSIDE	0	0
OUTSIDE	1	1

In this paper, new distance functions for R-Histograms were also introduced and shown to outperform the histogram intersection distance metric in experiments. We have also provided two fuzzy variants of prototype selection algorithms that demonstrate good experimental performance. Preliminary experiments showed very good annotation accuracy (92%–98% on synthetic images and 82%–93% on real images) with only 4–5 training images.

However, the system sometimes assigns a small value (less than 0.1) when zero is obviously the correct value. To overcome this problem, one possible solution is to add a postprocessing step to filter out these small values; however, this may result in filtering out small, but correct values. Another possibility is to use other classifiers, for example, Support Vector Machines, together with fuzzy k-NN, and to combine the outputs using classifier fusion methods.

The general framework presented here can also be used for automatic generation of other types of metadata. We are currently evaluating the system with more images. Our next goal is to investigate other fuzzy prototype selection algorithms and extend the system to support 3D medical images.

The R-Histogram method used here assumes that the objects are homeomorphic to a 2-ball, and it only considers pixels on the object boundary. For objects with complicated topology, we need to extend the current the R-Histogram method by taking into account all the pixels in the objects. We are currently developing efficient algorithms for the extended R-Histogram.

The current system takes segmented images as input. It is possible to incorporate automatic image segmentation algorithms, such as the Expectation-Maximization based algorithm used in Blobworld [5], into the system.

Table 12: Annotation accuracy on real images

Classifier	Accuracy
LEFT OF/RIGHT OF	0.89
ABOVE/BELOW	0.88
NEAR/FAR	0.82
INSIDE/OUTSIDE	0.93

6. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This work was supported in part by the National Science Foundation under grants ITR-0312629 and IDM-0083423.

7. REFERENCES

[1] PRTools3. `http://www.ph.tn.tudelft.nl/~bob/PRTOOLS.html`, 2003.

[2] M. Bereau and B. Dubuisson. A fuzzy extended k-nearest neighbors rule. *Fuzzy Sets and Systems*, 44(1):17–32, 1991.

[3] I. Bloch. Fuzzy relative position between objects in image processing: a morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(0162-8828):657–664, 1999.

[4] I. Bloch and A. Ralescu. Directional relative position between objects in image processing: a comparison between fuzzy approaches. *Pattern Recognition*, 36(7):1563–1582, 2003.

[5] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.

[6] H. L. Chieu and H. T. Ng. A maximum entropy approach to information extraction from semi-structured and free text. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pages 786–791, 2002.

[7] A. G. Cohn and S. M. Hazarika. Qualitative spatial representation and reasoning: an overview. *Fundamenta Informaticae*, 46(1-2):1–29, 2001.

[8] B. Dasarathy. *Nearest Neighbor Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, 1991.

[9] B. Dasarathy, J. Sanchez, and S. Townsend. Nearest neighbour editing and condensing tools–synergy exploitation. *Pattern Analysis and Applications*, 3(1):19–30, 2000.

[10] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the QBIC system. *Computer*, 28(9):23–32, 1995.

[11] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156171, 1975.

[12] G. Giuffrida, E. C. Shek, and J. Yang. Knowledge-based metadata extraction from postscript files. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, pages 77–84. ACM Press, 2000.

[13] V. N. Gudivada and V. V. Raghavan. Design and evaluation of algorithms for image retrieval by spatial similarity. *ACM Transactions on Information Systems*, 13(2):115–144, 1995.

[14] H. Han, C. L. Giles, E. Manavoglu, H. Zha, Z. Zhang, and E. A. Fox. Automatic document metadata extraction using support vector machines. In *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 37–48. IEEE Computer Society, 2003.

[15] P. Hart. The condensed nearest neighbor rule. *IEEE Transactions on Information Theory*, 14(3):515–516, 1968.

[16] T. Joachims. A statistical learning model of text classification with support vector machines. In *Proceedings of the 24th ACM SIGIR International Conference on Research and Development in Information Retrieval*, pages 128–136, 2001.

[17] A. Jóźwik. A learning scheme for a fuzzy k-nn rule. *Pattern Recognition Letters*, 1:287–289, 1983.

[18] J. M. Keller, M. R. Gray, and J. A. Givens. A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 15(4):580–585, 1985.

[19] R. Krishnapuram, J. Keller, and Y. Ma. Quantitative analysis of properties and spatial relations of fuzzy image regions. *IEEE Transactions on Fuzzy Systems*, 1(3):222–233, 1993.

[20] L. I. Kuncheva. *Fuzzy Classifier Design*. Studies in Fuzziness and Soft Computing. Physica-Verlag, Heidelberg; New York, 2000.

[21] W. Lam, C.-K. Keung, and D. Liu. Discovering useful concept prototypes for classification based on filtering and abstraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1075–1090, 2002.

[22] E. D. Liddy, E. Allen, S. Harwell, S. Corieri, O. Yilmazel, N. E. Ozgencil, A. Diekema, N. McCracken, J. Silverstein, and S. Sutton. Automatic metadata generation & evaluation. In *Proceedings of the 25th ACM SIGIR International Conference on Research and Development in Information Retrieval*, pages 401–402. ACM Press, 2002.

[23] P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera. Linguistic description of relative positions in images. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 31(4):573–588, 2001.

[24] P. Matsakis and L. Wendling. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):634–643, 1999.

[25] K. R. McKeown, N. Elhadad, and V. Hatzivassiloglou. Leveraging a common representation for personalized search and summarization in a medical digital library. In *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 159–170, 2003.

[26] K. Miyajima and A. Ralescu. Spatial organization in 2d images. In *Proceedings of the Third IEEE Conference on Fuzzy Systems*, volume 1, pages 100–105, 1994.

[27] K. Miyajima and A. Ralescu. Spatial-organization in 2d segmented images: Representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems*, 65(2-3):225–236, 1994.

[28] A. Pentland, R. W. Picard, and A. Sclaroff. Photobook: Content based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.

[29] D. B. Skalak. Prototype and feature selection by sampling and random mutation hill climbing algorithms. In *Proceedings of the 11th International Conference on Machine Learning*, pages 293–301, 1994.

[30] V.-W. Soo, C.-Y. Lee, C.-C. Li, S. L. Chen, and C. chih Chen. Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques. In *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 61–72, 2003.

[31] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[32] A. Takasu. Bibliographic attribute extraction from erroneous references based on a statistical model. In *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 49–60. IEEE Computer Society, 2003.

[33] J. Z. Wang and J. Li. Learning-based linguistic indexing of pictures with 2-D MHMMs. In *Proceedings of the 10th ACM International Conference on Multimedia*, pages 436–445. ACM Press, 2002.

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

[35] X. M. Wang and J. M. Keller. Human-based spatial relationship generalization through neural/fuzzy approaches. *Fuzzy Sets and Systems*, 101(1):5–20, 1999.

[36] Y. Wang and F. Makedon. R-histogram: Quantitative representation of spatialrelations for similarity-based image retrieval. In *The 11th Annual ACM International Conference on Multimedia*, pages 323–326, Berkeley, California, USA, 2003.

[37] D. L. Wilson. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man and Cybernetics*, 2(3):408–421, 1972.

[38] M.-S. Yang and C.-H. Chen. On the edited fuzzy-nearest neighbor rule. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 28(3):461–466, 1998.