Fuzzy Semantic Labeling for Image Retrieval

Margarita C. S. Paterno, Fun Siong Lim, Wee Kheng Leow
Dept. of Computer Science, National University of Singapore
3 Science Drive 2, Singapore 117543
paternom, limfs, leowwk@comp.nus.edu.sg

Abstract

This paper proposes a fuzzy image labeling method that assigns multiple semantic labels together with confidence measures to each region in an image. The confidence measures are derived from the distance of the region to hyperplanes constructed by support vector machines. Test results show that this method yields higher classification accuracy and retrieval precision than crisp labeling methods that are based on crisp classification.

1. Introduction

There are two general approaches to semantic labeling of image regions: crisp labeling and fuzzy labeling. Crisp labeling methods assign a single semantic label to each region. These methods determine a region's label by classifying the region's color and texture features using methods such as clustering, nearest neighbor classification and neural networks. In [1] and [2], one-vs-rest support vector machines (SVM) were used to perform crisp labeling but on images as a whole rather than on image regions. Due to the difficulty of accurately classifying an image region's features, these methods have been shown to work on at most 11 semantic classes [3, 4, 5, 6]. Fuzzy labeling method [7] assigns multiple semantic labels together with confidence measures to an image region. Although [7] illustrated the labeling of 30 semantic classes, it has not clearly demonstrated the strength of fuzzy labeling compared to crisp labeling in image retrieval.

This paper proposes another fuzzy labeling method that measures the confidence based on the orthogonal distance of an image region's feature vector to the hyperplane constructed by a support vector machine (SVM). The confidence measures assigned to a region represent the *signature* of the region and are used for region matching during image retrieval. Test results show clearly that our fuzzy labeling method yields higher classification accuracy and retrieval precision than crisp labeling based on crisp classification.

2. Algorithms

2.1. Semantic Labeling Methods

2.1.1. Crisp Labeling. One way to label image regions using crisp labels is to classify a region i using a multi-class classifier, such as neural network and the Directed Acyclic Graph (DAG) SVMs [8], into one of m semantic classes, say, class c. Then the crisp label of region i is simply c.

Another crisp labeling method involves training m one-vs-rest binary SVMs [1, 2, 9] where the jth SVM is trained to classify regions into either class j or non-class j. After training, a region i is classified using all the m one-vs-rest SVMs. Then region i is assigned the crisp label c if, among the SVMs that classify region i positively, the cth SVM returns the longest distance between region i's feature vector and its hyperplane. If no SVM classifies region i as positive, then region i is labeled as "unknown".

2.1.2. Fuzzy Labeling. Similar to the second crisp labeling method, our proposed fuzzy labeling method also trains m one-vs-rest binary SVMs on training samples. After training, a *confidence curve* is approximated for each SVM as follows. Validation samples are classified by an SVM and their distances to the hyperplane are computed. The samples are sorted in increasing order and an algorithm [7] is applied to recursively partition the range of distances into intervals such that the classification accuracy within each interval can be measured and the accuracy changes smoothly from one interval to the next.

Figure 1 shows a sample confidence curve constructed. Given a test sample, its distance to the SVM's hyperplane is computed and its expected classification accuracy is obtained from the confidence curve using linear interpolation and regarded as the confidence measure. Since the classification accuracy is bounded between 0 and 1, the confidence curves also provide nonlinear normalization of the distances to confidence measures within the [0, 1] range. Now the

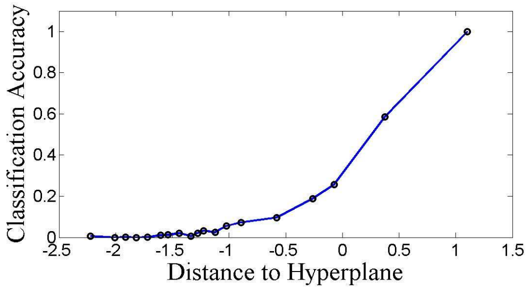

Figure 1. A sample confidence curve.

test sample can be assigned a fuzzy label or *signature* in the form of a vector

$$\mathbf{v} = [v_1\ v_2 \ldots v_m]$$

where v_j is the confidence that the sample belongs to class j.

With the first crisp labeling method, a sample's signature would be $\mathbf{v}$ such that exactly one of the v_j's is 1 while the rest are 0. With the second crisp labeling method, at most one of the v_j's is 1, and the signature of an "unknown" region would be a zero vector.

2.2. Region Matching

During image retrieval, a query image is constructed using regions of known semantic classes. So, image matching is performed by matching regions in the database images with known regions in the query image. The signatures of the known regions are computed as follows. For each semantic class c, k-means clustering is performed on the validation samples in class c according to their signatures. The appropriate number of clusters k is chosen using silhouette values [10] that measure how well the samples are clustered. The prototype of each cluster, i.e., the mean signature of the samples in the cluster, serves as a prototype signature of the semantic class. Thus, each semantic class can have more than one prototype label. Having multiple prototype signatures $\mathbf{p}_k$ improves the retrieval performance because there is a large variation of signatures even within a single semantic class.

In empirical tests, it was found that prototype signatures $\mathbf{p}_k$ of a semantic class c where $\max_j\{p_{kj}\} \neq p_{kc}$ are misleading. That is, the signature of class c says that the confidence of belonging to class c is lower than those of other classes. These prototype signatures are *unreliable* and should not be used in region matching.

For crisp labeling, each semantic class has only one prototype signature. The prototype signature $\mathbf{p}$ of class c is such that $p_c = 1$ and $p_j = 0$ for all $j \neq c$.

Given the signature $\mathbf{v}$ of a region r and prototype signatures $\mathbf{p}_k$ of a class c, the distance d between the region and the class is simply the minimum Euclidean distance between $\mathbf{v}$ and $\mathbf{p}_k$:

$$d(r,c) = \min_k d(\mathbf{v}, \mathbf{p}_k)\ . \qquad (1)$$

2.3. Image Matching

A query image Q contains n known regions R_i each having several prototype signatures $\mathbf{p}_{ik}$. A database image I contains many regions r_l with signatures $\mathbf{v}_l$. Two kinds of distances between query and database image can be defined. The first matches regions in Q with regions in I at the *same* positions. Denote $r_i \in I$ as the region at the same position as $R_i \in Q$. Then

$$d(Q,I) = \frac{1}{n}\sum_i d(R_i, r_i)\ . \qquad (2)$$

The second kind matches regions in Q with regions in I at *any* position. Then $d(Q, I)$ should be defined in terms of the best matching regions between Q and I:

$$d(Q,I) = \frac{1}{n}\sum_i \min_l d(R_i, r_l)\ . \qquad (3)$$

3. Performance Test

3.1. Image Data Sets

A variety of 31 semantic classes shown in Figure 2 were identified. For each semantic class, 550 image blocks of size 64×64 pixels that contained objects of a single semantic class were cropped from images in the Corel library of 50,000 photos. Of these, 375 image blocks were chosen at random to form the training set, 125 for the validation set and 50 for the test set.

In order to assess how well the labeling method can generalize to image blocks that are not well-cropped, an additional 800 images were selected from the Corel photo library to form a general test set. Each semantic class was contained in at least 25 images. Each image was partitioned at regular intervals into 77 overlapping image blocks of size 64×64 pixels. Each image block was manually assigned a label to denote the ground truth, which was either one of the 31 semantic classes, "unknown" if it did not belong to any of the 31 semantic classes, or "ambiguous" if it contained objects in more than one semantic class. In total, there were 26179 (42.5%) blocks with known classes, 4588 (7.4%) unknown blocks and 30833 (50.1%) ambiguous blocks.

Four different types of low-level features, namely, fixed color histograms, Gabor features, multiresolution simultaneous autoregressive features (MRSAR) and edge histograms, were extracted from each image block, normalized and combined to form a feature vector of 274 dimensions.

3.2. Region Classification Test

To assess the accuracy of fuzzy labeling quantitatively, a region classification test was performed. The distance $d(r, c)$ (Eq. 1) between sample r and class c is computed and sample r is classified to the nearest class in terms of $d(r, c)$. Four labeling methods were compared: crisp labeling methods using DAG SVM and one-vs-rest (1vsR) SVMs, and fuzzy labeling using all prototype signatures and reliable prototype signatures only. The tests were performed on both the well-cropped and general test samples.

Test results in Table 1 show that all the methods except for crisp labeling with one-vs-rest SVMs can label all test image blocks. The latter method can label only 19.1% of the general test samples.

For the well-cropped test samples, fuzzy labeling with reliable prototype signatures achieves the highest classification accuracy followed closely by crisp labeling with DAG SVM. For the general test samples, the accuracy of both crisp labeling methods drops drastically to 10%. This indicates that the general test samples are much harder to classify, probably due to more noise and ambiguity in the samples. In spite of this, both fuzzy labeling methods outperform both crisp labeling methods. In particular, fuzzy labeling with reliable prototype signatures achieves a classification accuracy of 24.7% on blocks with *known* classes, which is more than twice the accuracy of crisp labeling.

3.3. Image Retrieval Test

For image retrieval test, 31 query images were constructed by placing a region of a known semantic class at an appropriate position in each query image. Then the 800 database images were compared with each query image and ranked in order of increasing distance. Since it is laborious to manually identify the relevant images among all 800 images for each of the 31 queries, retrieval precision was measured only among the first N retrieved images and averaged over the 31 queries. This retrieval test was performed separately for same-position (Eq. 2) and any-position (Eq. 3) retrieval using each of the crisp labeling and fuzzy labeling methods.

Figure 3 shows that retrievals using fuzzy labeling perform better than those using crisp labeling, especially for same-position retrieval. Moreover, fuzzy labeling with reliable prototype signatures achieves the best retrieval performance which is consistent with the results for region classification test. Note that retrieval precision with reliable prototype signatures ranges from 26% to 37% for same-position retrieval and 35% to 49% for any-position retrieval. This shows that the method can still retrieve many relevant images even though its region classification accuracy is only 24.7%.

4. Conclusion

This paper proposes a fuzzy image labeling method that assigns multiple semantic labels together with confidence measures to an image region. This is done by training m one-vs-rest classifiers and measuring the classification confidence with confidence curves derived using validation samples. Prototype fuzzy signatures for each semantic class are obtained by applying k-means clustering on the validation samples. Image retrieval is then performed by matching the signatures of the region in the image with the prototype signatures of the known regions in a query image.

Experimental tests show that fuzzy labeling performs better than crisp labeling in semantic region classification and in semantic image retrieval especially for same-position image retrieval.

References

[1] G. Wu, E. Chang and C. Li, "BPMs versus SVMs for image classification", In *Proc. IEEE Intl. Conf. on Multimedia*, 2002, pp. 505-508.

[2] K. Goh, E. Chang and T. Cheng, "SVM binary classifier ensembles for image classification", In *Proc. ACM Conf. Info. and Knowledge Mgt.*, 2001, pp. 395-402.

[3] N. W. Campbell, W. Mackeown, B. T. Thomas and T. Troscianko, "Interpreting image databases by region classification", *Pattern Recognition*, 30, 1997, pp. 555-563.

[4] M. Szummer and R. W. Picard, "Indoor-outdoor image classification", In *Proc. ICCV Workshop on Content-based Access of Image and Video Databases*, 1998, pp. 42-51.

[5] C. Y. Fung and K. F. Loe, "Learning primitive and scene semantics of images for classification and retrieval", In *Proc. ACM Multimedia*, 1999, pp. II: 9-12.

[6] C. P. Town and D. Sinclair, Content Based Image Retrieval Using Semantic Visual Categories, Technical Report 2000.14, AT&T Laboratories, Cambridge, 2000.

[7] R. Li and W. K. Leow, "From Region Features to Semantic Labels: A Probabilistic Approach", *Proc. Int. Conf. on Multimedia Modeling*, 2003, pp. 402-420.

[8] J. Platt, N. Cristianini and J. Shawe-Taylor, "Large margin DAGs for multiclass classification", *Advances in Neural Information Processing Systems*, 12, 2000, pp. 547-553.

[9] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

[10] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, *J. of Comp. and Appl. Mathematics*, 20, 1987, pp. 53–65.

Figure 2. Sample images of 31 semantic classes used.

Table 1. Test results on (a) well-cropped test samples and (b) general test samples.

	(a) well-cropped test samples				(b) general test samples			
	Crisp labeling		Fuzzy labeling		Crisp labeling		Fuzzy labeling	
	DAG	1vsR	All	Reliable	DAG	1vsR	All	Reliable
% labeled	100.0%	70.6%	100.0%	100.0%	100.0%	19.1%	100.0%	100.0%
accuracy	58.3%	51.2%	52.5%	**60.8%**	10.4%	10.5%	19.8%	**24.7%**

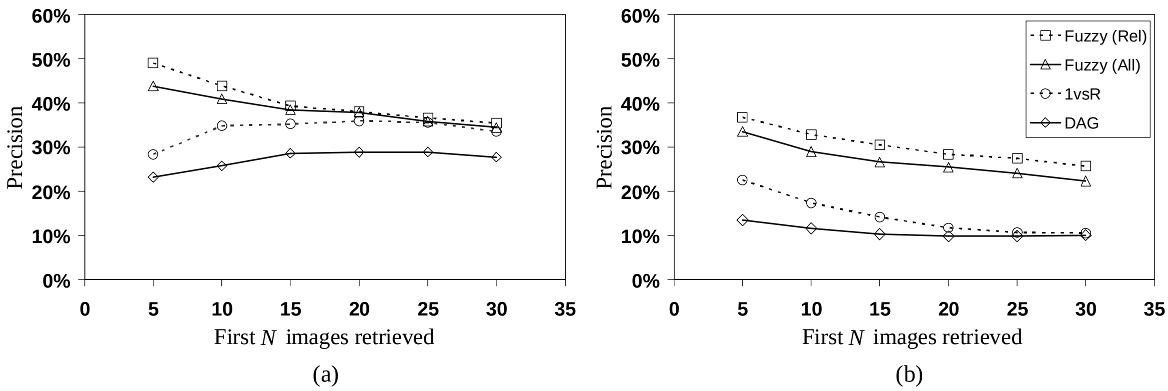

Figure 3. Average precision over all classes for (a) any-position and (b) same-position retrieval.