MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Probabilistically Semantic Labeling of IR Image for UAV

Teng Li , Jihwan Woo
Department of Electrical Engineering,
KAIST 373-1 Guseong-dong, Yuseong-gu,
Daejeon, Korea
tengli, jhwoo@rcv.kaist.ac.kr

In So Kweon
Department of Electrical Engineering,
KAIST 373-1 Guseong-dong, Yuseong-gu,
Daejeon, Korea
iskweon@kaist.ac.kr

Abstract

Applying computer vision technology to IR (Infra-Red) images for UAV (Unmanned Aerial Vehicle) applications is difficult due to its characteristics which differ from common image processing. By combining visual categorization with low level IR image processing, this paper presents a framework for automatic labeling of IR images in probabilistic manner. We extract the features which contain temperature, texture and orientation information from the IR image, model visual categories by the distribution of features in terms of an extended visual vocabulary, and categorize IR image segments probabilistically. The proposed framework is demonstrated in experiments with high labeling accuracy, for near IR images of urban terrain taken from 100 feet altitude.

1. Introduction

Due to the increasing requirement of UAV applications such as traffic surveillance, environment monitoring, etc, many researchers have tried to improve autonomous flight capability of UAV in the last decade [1]. Computer vision technology is applied and the UAV needs to carry out tasks like obstacle detection, object tracking and motion estimation for its navigation [2].

Previous works on UAV have used visible color images but the information of visible color images can be unavailable when UAV navigates under weak illuminated conditions. Compared with visible color images, IR images contain more information in the darkness with its own advantages. But IR images have low resolution and they are noisier than color images thus it is difficult to apply computer vision algorithms directly [2].

In this paper, we propose an autonomous probabilistic semantic labeling system for aerial IR images captured in urban terrain. This system segments the IR image to semantically meaningful regions and probabilistically labels the regions with object categories. The resulted labeled images can be applied to safe landing place detection, building detection and road tracking. Also we can incorporate prior semantic knowledge easily based on this and take further operations.

Low level or middle level image processing alone is not enough to realize robust semantic labeling for IR images. In this paper, we combine high level visual categorization with low level image processing and image segmentation for this task.

Visual categorization is a difficult but hot problem in computer vision. In recent years, many approaches for this topic have been proposed, Several algorithms are built around a vocabulary of visual terms and model visual categories with the histogram of visual word count [3-6]. In this paper, we model and categorize segments of the IR image based on such approach. However, what distinguishes our categorization from previous works is our design of features for IR image processing, combined models based on two types of features, and finally, the prediction of probabilities that the image region might contain categories. We use SVM classifier in probabilistic form for this objective.

2. System framework

Figure 1 shows the overall structure of the proposed semantic labeling system for IR images. In the training process, we extract features and learn visual category models from the manually segmented and labeled training images. Then given an IR image, we automatically segment it to regions, extract the features of each region, and categorize these regions based on the learned model. This process does not require precise segmentation. Unlike the common object categorization tasks, we do not predict a single class label for each image region, but predict the probabilities that it may be composed of categories. Finally, we get a probabilistically semantic labeling graph of the IR image.

In section 3 and section 4 we will present more detailed process of feature extraction, visual category model training and probabilistic semantic labeling for IR images.

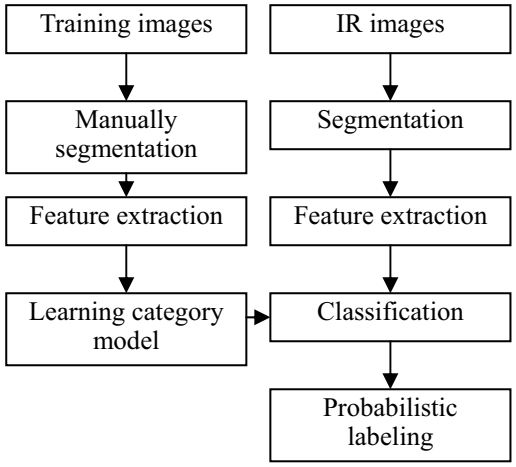


Figure 1. Overview of the whole system process

3. Visual category model training

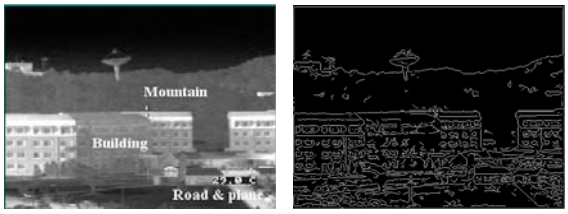


Figure 2. An aerial IR image and its Edge map

3.1. Feature extraction

Figure 2 shows an example of manually semantic labeling of an aerial IR image using categories as building, road or plane, and mountain. The right column shows its edge map. We can observe that the intensity value and local edge map characterize different categories well.

Intensity value of each pixel in IR images shows the temperature radiation. And we test a number of different filter-banks made of Gaussians, first and second order derivatives of Gaussians and Gabor kernels for analyzing texture and boundary, as a kind of 'texton' feature. Similar 'texton' feature has been used in some of the previous works [4].

The filter-banks of texton feature are made of 3 Gaussians (with σ=1, 2, 4), 4 Laplacian of Gaussians (LoG) (with σ=1, 2, 4, 8) and 4 first order derivatives of Gaussinans (with σ=2, 4 and into x, y directions). Therefore, each pixel in each image is associated with an 11-dimensional texton feature vector.

The statistics of pixel intensity value in an image region contains the global distribution information of temperature in the region. While texton feature is relate to the texture and linearity of object boundary in the region. Combination of the two features can help to discriminate confusing regions. It is useful for categorizing automatically segmented regions, which usually do not contain a pure object category.

3.2. Visual categories modeling

To model the categories of image segments, we use the histogram of features in the segment based on a visual vocabulary, with an assumption that visual codes distributions of a given class are similar. The essential idea of this approach is to provide an intermediate representation which helps to bridge the semantic gap between the low level features extracted from an image and the high level classification algorithms [3].

Figure 3 shows the principles and process of our visual category modeling. Given the training set, after extracting features, learn a visual vocabulary by clustering training features to some visual codes. By assigning features of images to the vocabulary with a vector quantization algorithm, and counting the number of features assigned to each code, we can get a distribution (histogram) over the vocabulary for each image. Then we apply a multi-class classifier to the histogram representation of images, train category models and predict the probability that the category labels might be assigned to images.

To make full use of the features, we do not add intensity as simply another dimension feature to the texton feature, but use different vocabularies for two kinds of features. We combine the histograms of the two vocabularies for modeling object categories.

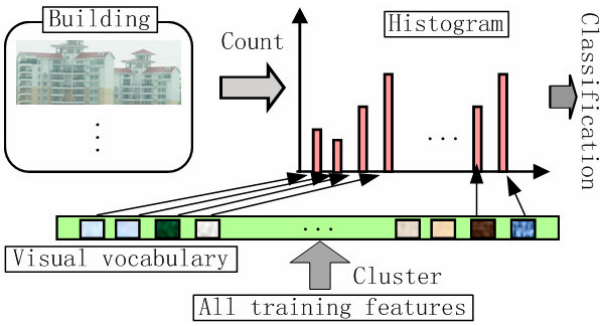


Figure 3. Visual vocabulary based category model

3.3. Visual vocabulary learning

Several algorithms aiming at constructing an efficient visual vocabulary have been proposed [4-6], usually based on the clustering or vector quantization algorithms.

We adopted a simple and efficient k-means clustering to learn the visual vocabulary of texton feature. More advanced clustering methods exist, but for visual categorization task in terms of the vocabulary, these algorithms give similar performance. Through experiments, the size of vocabulary for texton feature is set at 1000.

For intensity feature, we use 256 intensity values to compose the vocabulary.

Thus, each image region can be represented by a 1256 bin histogram in terms of this combined vocabulary.

4. Probabilistic semantic labeling

4.1. Segmentation

To segment IR images to regions that are semantically meaningful, we use K-means cluster algorithm. The input image is grouped into 4 regions by using intensity value. We apply the categorization algorithms for the resulted regions with grouped pixels. Each pixel has the intensity and texton feature.

Figure 5 displays an example of the segmentation result. Most resulted regions contain mainly one or two categories that high accuracy labeling is possible.

4.2. Classification and probabilistic labeling

After assigning features to the clusters of visual vocabulary and getting the histogram representation, we apply the mutli-class training and classification algorithms directly. We use Naïve Bayes classifier for its simplicity and high speed, and Support Vector Machine for its excellent performance in many classification problems.

4.2.1. Naïve Bayes classifier

Given a set of labeled image regions $R=\{R_i\}$ and a visual vocabulary $V=\{V_i\}$. Each pixel in a region is assigned to a code of the vocabulary to which it lies the closest. Write the number of times code V_t occurs in region R_i as $N(t,i)$. Then calculate the predict score that region R_i belongs to category C_j:

$$P(C_j \mid R_i) \propto P(C_j)P(R_i \mid C_j) = P(C_j)\prod_{h=1}^{|V|} P(V_t \mid C_j)^{N(t,i)} \quad (1)$$

The conditional probabilities of code V_t given

category C_j are computed according to the formula below, with Laplace smoothing:

$$P(V_t \mid C_j) = \frac{1 + \sum_{\{R_i \in C_j\}} N(t,i)}{|V| + \sum_{s=1}^{|V|} \sum_{\{R_i \in C_j\}} N(s,i)} \quad (2)$$

4.2.2. Support Vector Machine (SVM)

We adopt one-against-one strategy to apply SVM classifier for multi-class classification problem. 'RBF' is chosen as the kernel type. In our application to IR image, 'RBF' kernel shows good performance.

Originally SVM predicts only class label but not probability information. For our study, we use a method of producing probabilistic outputs proposed by Wu, et al [7].

Given k classes of data, for any x, the goal is to estimate

$$p_i = p(y = i \mid x), i = 1,...,k. \quad (3)$$

Firstly estimated pairwise class probabilities

$$r_{ij} = p(y = i \mid y = i \text{ or } j, x), i = 1,...,k. \quad (4)$$

using an improved implementation:

$$r_{ij} \approx \frac{1}{1 + e^{A\hat{f} + B}} \quad (5)$$

Where A and B are estimated by minimizing the negative log-likelihood function using known training data and their decision values $\hat{f}$. Obtain p_i from all these r_{ij} s by solving the following optimization problem (For the details of this process, please refer to [7]):

$$\min_P \frac{1}{2} \sum_{i=1}^{k} \sum_{j:j \neq i} (r_{ji} p_i - r_{ij} p_j)^2 \quad \text{subject to} \sum_{i=1}^{k} p_i = 1, p_i \geq 0, \forall i \quad (6)$$

4.3. Probabilistic labeling

Since segmentation algorithms can not perfectly segment images to semantic regions, there is usually a mix of categories in a region. We predict the probabilities that a segmented region may contain object categories, and label the region with probability of categories. Only the main categories whose relative probability is above a threshold are considered as meaningful.

5. Experiments and Discussion

The dataset used in experiments contains 500 near IR images of urban terrain captured from 100 feet altitude. We randomly select 47 images and precisely mark four kinds of object category regions: road or plane (R&P), building, water and mountain. We mark and label 30 regions for each category for training. To test our categorization algorithm, we coarsely mark 17 regions in other images and label as a validation set.

To evaluate the performance of our labeling system, we label 108 automatically segmented regions from 27 images, which are randomly selected and different from the training images. Segmented regions mainly composed of one category are labeled with one class no., and regions which are not pure are labeled with two candidate: the first class no. and the second.

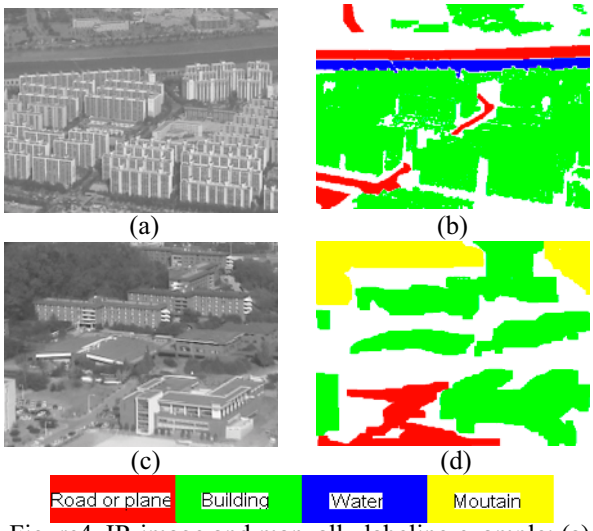


Figure4. IR image and manually labeling example: (a), (c): original images (the size is 320×240); (b), (d): the precise labeling of (a) and (c) respectively for training (best view in color).

Table 1 gives the comparison of our feature for IR images with SIFT and GRIF descriptors, which are popular in object categorization [8, 9]. We densely sample 10×10 patches in the image, sampling interval is set as 5, and compute SIFT and GRIF descriptors for each patch. Patches which have more than 60% pixels in a marked region are considered as belonging to this region. We build vocabularies of size 1000 for SIFT and GRIF respectively, and categorize feature histograms of regions based on the vocabularies. We get the 5-fold cross validation rate on training set. Obviously our feature set is more suitable here.

The last row of table 1 also gives the result of classifying individual features of categories directly, using SVM classifier. We can see the effect of our region modeling method by comparison.

Table 2 shows the SVM classification accuracy rates when we use intensity feature only, texton feature only and both. We evaluate the accuracy rates by three methods: five fold cross validation rate of training set, classification accuracy on validation set and test set. Here for test set, we only consider the first candidate label of regions. The combined feature is better than any single feature alone, especially on the test set. Intensity is good for pure marked regions, but not robust for mixed regions.

Note that some automatically segmented regions contain two main categories, and we have two candidate labels for these test regions. In table 3, we give the accuracy rates with assumption that resulting in the second label candidate is also correct. We compare the performance of Naïve Bayes and SVM, using the combined feature.

Table 4 gives the confusion matrix of classification of segmented regions when only the first candidate label is considered, using SVM and the combined feature.

To evaluate our probabilities output, we manually mark three regions for each category. Table 6 shows the output probabilities of these regions through different classifiers.

Figure 5 shows the segmentation result and in table 5 we give the probabilistic labeling result from SVM and Naïve Bayes classifier. There are four candidate labels

for each segment: Road or Plane (R&P), Building, Water and Mountain. We only label the candidates whose probability is above 20%.

The probabilistic output depends on the learned model, i.e., model parameters and training data affect the result. Our output can roughly represent the relative proportions of contents in the region, and we can adjust according to different application situation.

The classification performance of SVM classifier is a little better than Naïve Bayes classifier. However, Naïve Bayes is more effective considering speed. They provide two good choices for further works based on the proposed framework.

6. Conclusions and Future Work

In this paper, we have presented a framework for automatically probabilistic labeling of IR image for UAV. To this aim, we extracted features which characterize IR image regions well and incorporate the up to date visual categorization ideas. The proposed system gave high accuracy rates in experiments. The resulted semantically labeled IR images can be useful for diverse UAV applications.

In the near future, we will make a model which incorporates the spatial context information of the IR image and improve a matching algorithm by using these semantic labeling. Because finding point matching is difficult in IR image, the semantic labeling system proposed is expected to expand the domain of IR applications.

References

[1] Anibal Ollero, Joaquin Ferruz, Fernando Caballero, "Motion compensation and object detection for autonomous helicopter visual navigation in the COMETS system", *ICRA*, 2004.

[2] Jihwan Woo, et al, "Robust Horizon and Peak Extraction for Vision-based Navigation", *IAPR Machine Vision Application*, 2005.

[3] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, Cédric Bray, "Visual Categorization with Bags of Keypoints", *SLCV Workshop, ECCV,* 2004.

[4] J. Winn, A. Criminisi and T. Minka, "Object Categorization by Learned Universal Visual Dictionary", *ICCV*, 2005.

[5] Sungho Kim, In So Kweon, "Simultaneous Classification and Visual Word Selection using Entropy-based Minimum Description Length", *ICPR*, Hong Kong, 2006.

[6] Frederic Jurie and Bill Triggs, "Creating Efficient Codebooks for Visual Recognition", *ICCV*, 2005.

[7] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling", *Journal of Machine Learning Research*, 5:975–1005, 2004.

[8] Lowe, D.G., "Distinctive image features from scale-invariant keypoints", *IJCV* 60(2004) 91−110.

[9] Sungho Kim and In So Kweon, "Biologically Motivated Perceptual Feature: Generalized Robust Invariant Feature", *ACCV*, 2006.

Table 1. Performance comparison using different feature sets and model methods on training set.

	SIFT	GRIF	Tex.	Inten.+Tex.
Histogram	77.9%	81.1%	84.2%	95.8%
Single	56.2%	56%	47%	×

Table 2. Classification result of different features.

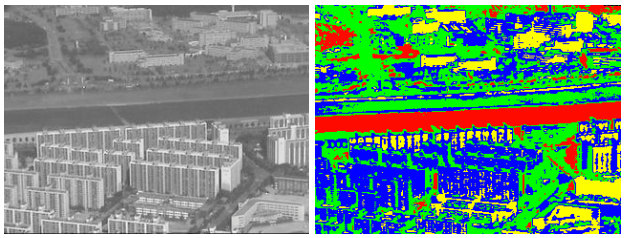
Feature	SVM classification accuracy		
	5-fold c.v.	Validation set	Test set
Intensity	95%	88.2%	70.4%
Texton	84.2%	82.4%	58.3%
Intensity+Texton	95.8%	88.2%	83.3%

Table 3. Recognition result.

Classifier	Classification accuracy	
	1 candidate	2 candidates
Naïve	81.5%	93.5%
SVM	83.3%	94.4%

Table 4. Confusion matrix for 4 categories (1 candidate).

True label	Inferred label			
	R&P	Building	Water	Mountain
R&P	22	9	0	0
Building	3	41	0	0
Water	0	0	21	0
Mountain	6	0	0	6



(a) (b)

Figure 5. Automatic segmentation result. (a): an IR image; (b): segmentation result of (a)(best view in color). Corresponding probabilistic labeling is given in table 5.

Table 5. Probabilistic labeling for image segments: 'R', 'B', 'G' and 'Y' correspond to red, blue, green and yellow segments in figure 5 respectively.

	SVM		Naïve	
Segment	1st label	2nd label	1st label	2nd label
R	Water 45.5%	R&P 37.9%	Water 74.1%	Building 27.7%
G	R&P 95%	×	R&P 94.1%	×
B	Building 94%	×	Building 69%	R&P 30.9%
Y	Building 97%	×	Building 91.3%	×

Table 6. Average output probabilities of different categories for manually marked regions.

Classifier	R&P	Building	Water	Mountain
SVM	88%	93.7%	84.9%	90.7%
Naïve	83.8%	88.5%×	89.7%	97%×