Robust Object Recognition Using Color Co-occurrence Histogram and Spatial Relations of Image Patches

Heebeom Bang, Sanghoon Lee, Dongjin Yu, and Il Hong Suh
Dept. of Electrical and Computer Engineering
Hanyang University
Seoul, Korea, 133-791

Abstract

In this paper, a robust object recognition system is proposed, where patch-based pyramid images and spatial relationships among patches are utilized for our object model. Specially, both color histogram (CH) and color co-occurrence histogram (CCH) are applied to obtain image features for each patch. Locations of candidate regions to be tested are decided by a particle filter in our matching process. We show that the performance of object recognition can be improved by using spatial relationships among patches. To show the validities of our proposed method, there are employed input images from various environments as test images.

1 Introduction

The object recognition system has to be designed to be dependable especially for a service robot. Object recognition is not an easy task, since there are changes of illumination, object size, orientation and viewpoint. For these reasons, it is essential to devise both modeling and matching processes in such a way that the processes are robust to changes of environmemtal conditions.

Many researchers have proposed novel models to be robust to the changes of environmental conditions. Chang proposed a model using a color co-occurrence histogram. This model contains not only a color histogram but also geometric information as a feature. In addition, the model was built by images taken by various viewpoints to build a model to be invariant for rotation. Although his method was somewhat invariant to rotation, it was not reliable for scale changes [1]. Selection of a color model is crucial for color-based object recognition. Redfield utilized only 16 colors to describe an object and implemented the matching process with those colors. He assumed that illumination condition was fixed. This might cause his method an reliable for changes of environmental conditions [2]. Zhang proposed object recognition system based on multi-scale affine invariant image regions [11]. On the other hand, part(patch)-based model has been applied to build the model robust to partial occlusions, where relationships among patches are utilized to solve partial occlusions. It has been reported that the recognition performance could be improved by using those patch-based models [3] [6] [8] [12] [13].

There are some research works focusing on developing robust matching methods, instead of proposing novel models. They proposed learning-based object recognition methods for better performance [4] [7] [9]. Learning from various environments can improve the performance, but it needs lots of time to ensure reliable learning performances.

In this paper, we will propose processes of both object modeling and object matching to be robust to changes of environmental conditions as much as possible. The changes of illumination and rotation and partial occlusions are examples for those environmental changes. First, pyramid images are utilized to build a robust model for scale-invariance. Pyramid images are divided into patches to deal with partial occlusions. The performance of object recognition is improved by adding spatial relationships among patches to a model. Both color histogram and color co-occurrence histogram are used as features for each patch. And, there are employed representative colors in HSV color space. On the other hand, for the matching, locations of candiate regions similar to reference image patches are found by a prticle filter in the whole input images. Our system recognizes the object by comparing condidate regions and patches. Moreover, spatial relationships among patches and condidate regions are also included for our matching process.

To show the validities of our proposed method, there are employed input images from various environments as test images.

2 Overview of proposed robust object recognition system

Object recognition methods can be classified by a top-down approach and a bottom-up approach. In a top-down approach, we find a query object in the input image. In a bottom-up approach, we examine what objects are included in the input image. In this work, object recognition is implemented by a top-down approach. We build object models first, and find out whether the query object is in the input image or not. Then, we finally identify the highly probable location of regions similar to these models in the input image.

2.1 Process of object modeling

Fig. 1 shows the process of building an object model. First, we take object images from each viewpoint. Second, we transform an object image taken by each viewpoint into image pyramid. In during quantization, we transform RGB colors into quantized colors in HSV color space. Fourth, we divide these images into regions which have the same sizes. Those regions we take from this step are called patches. Consequently, each scaled image is divided into a number of patches. For each patch, the information of viewpoint, scale level, position, color histogram (CH) and color co-occurrence histogram (CCH) are contained.

2.2 Match Process

A matching process is started from taking a query and an input image. The query is utilized to call a model in our database. When a matching process is finished, we can find out the exact location of the query object in the input image.

Block diagram of our matching process is shown in Fig. 2. First, we choose a specific queried model from our database. After that, we have to select a candidate region in the input image. Candidate region is a certain area in the input image, which has to be closely matched with a patch of an object model. As a whole input image cannot be matched with a single patch, we should select a candidate region for the patch in the input image. The important thing is that locations of those candidate regions needs to be correctly estimated. For this, we utilize a particle filter. Candidate regions in the input image are selected by a particle filter. In the matching process, features from those candidate regions are compared with the features of the query object. Among those candidate regions, the candidate regions which have similar feature values with the specific patches are selected. For those selected candidate regions, we examine the spatial relationships among their neighboring candidate regions. Consequently, the location of the query object in the input image is finally determined by matching a feature value and by comparing the spatial relationships among neighboring candidate regions.

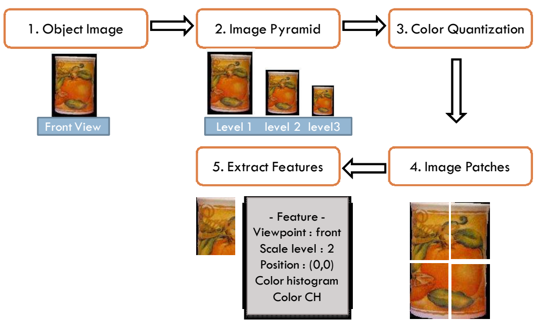

Figure 1: Flow chart of object modeling process

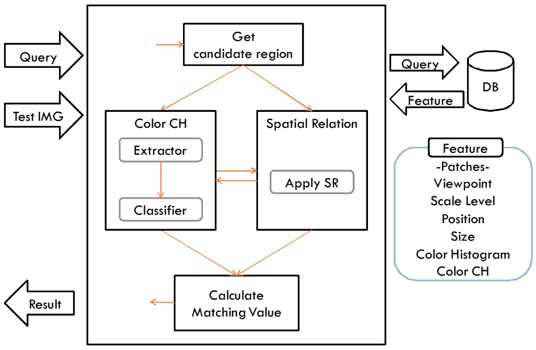

Figure 2: Block diagram of matching process

3 Particle Filter

Particle Filter is used to search for highly propable candidate regions where parts of the query object can be found. As we use the spatial relationships among candidate regions, we can get the highly propable candidate regions crucially.

Particle filter is a kind of Bayesian filter which computes posterior distribution by using several weighted samples. Bayesian filter consists of 2 steps: Prediction and Update. In the prediction step, probabilistic system transition model $p(x_t \mid z_{1:t-1})$ is used in given observation $z_{1:t-1}$ to predict the posterior at time t.

$$p(x_t \mid z_{1:t-1}) = \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid z_{1:t-1}) dx_{t-1} \tag{1}$$

If we measure z_t at time t, the state can be updated by applying Bayes' Rule.

$$p(x_t \mid z_{1:t}) = \frac{p(z_t \mid x_t) p(x_t \mid z_{1:t-1})}{p(z_t \mid z_{1:t-1})} \tag{2}$$

In a particle filter, the posterior $p(x_t \mid z_{1:t})$ is approximated by a finite set of N samples $\{x_t^i\}_{i=1,..,N}$ with importance weights w_t^i. Sample x_t^i are drawn from the weight of the samples. The weights become the observation likelihood $p(z_t \mid x_t)$.

The weight of a particle is obtained from comparing the patch in object model with the color histogram of the particle. As the computational burden for color co-occurrence histogram is heavy, the color histogram is applied instead. Observation z_t stands for the color histogram of a particle in the input image.

To reduce computational burden, sizes of all patches are fixed to be the same. We can find candidate regions efficiently through this method, because all particles have the same sizes. If the sizes of patches are different, the size of particles must be different according to those of patches. In addition, the value of color histogram from each particle can be reusable.

The major issues for a particle filter are how to draw initial particles and how to implement a re-sampling process. In this paper, we draw initial particles uniformly and remove the particles which have low weight values for a re-sampling process. The particles left in the end will be chosen as candidate regions. A matching process is implemented by color co-occurrence histogram for those candidate regions.

4 Color Co-occurrence Histogram

4.1 Color Model

There are various color models. Each color model has its own characteristic. HSV color model consists of three components, Hue, Saturation and Intensity. The value of Hue is generally defined from 0 degree to 360 degree. Both the value of Saturation and Intensity are defined from 0 to 1. The advantage of HSV color space is that most meaningful color demarcations (chromatic/achromatic; light/dark; dark/color/black; light color/white) can be made with simple threshold operations. Here, we adopt a tree quantization method in HSV [2]. The color pixel C(R, G, B) which is defined in RGB color space can be defined as a representative color(C(red), C(black),etc.).

4.2 Color Co-occurrence Histogram

We use color co-occurrence histogram as the feature of a patch. The Color CH holds the number of occurrences of pairs of color pixels C1= (red) and C2= (black) separated by a vector in the image plane $(\Delta x, \Delta y)$. We can write the color co-occurrence histogram symbolically as $CCH(C1, C2, \Delta x, \Delta y)$. In order to make CCHs invariant to rotation in the image plane, we ignore the direction of $(\Delta x, \Delta y)$ and keep track of only the magnitude d given by

$$d = \sqrt{\Delta x)^2 + (\Delta y)^2}. \tag{3}$$

We can rewrite the co-occurrence histogram symbolically as $CCH(C1, C2, d)$. In Color CH, we have to define the set of distance (D) and the set of Color (C). Here, as the set of color, C, fifteen representative colors are employed, and distance, D is discretized as one of twelve integer values as in [1].

4.2.1 Similarity

Similarity between a patch of query object and a candidate region in the input image can be computed as

$$S(r,q) = \eta \sum_{c_1=1}^{n_c} \sum_{c_2=c_1}^{n_c} \sum_{d=1}^{n_d} \min(CCH_r(c_1,c_2,d), CCH_q(c_1,c_2,d)), \tag{4}$$

where r, q, n_d, η are defined as follows;
S: similarity
r: candidate region of the input image
q: patch of query object
n_c: number of color set
n_d: number of distance set
η: normalization factor

5 Spatial Relation

5.1 Modeling Process

Object models are divided into patches in their different scale levels. Each patch has one feature independently. Consequently, we have a multi-feature map. However, a single patch cannot describe a single object. A set of patches may be required to describe a single object. Here, we consider the spatial relationships among patches by using their geometry relations.

Figure 3: Five viewpoints of object model (Top, Left, Right, Front, Back)

To take out model patches from a reference object image, five different viewpoints are considered as shown in Fig. 3, where view from bottom is excluded. The five corresponding images are obtained as in Fig. 3. Image for each view is transformed into various scales. After scale transformation, those scaled images are divided into patches of the same size. Consequently, we can have many patches. By using geometric relationships among those patches, matching performance can be efficiently enhanced.

5.2 Matching Process

Spatial relation in matching process is applied as follows. First, $\mathbf{R}^{\max}$ is defined as the set of candidate regions. Each element of this set has the highest similarity for each patch of query object. $\mathbf{R}^{\max}$ can be obtained by

$$\mathbf{R}^{\max} = \{r_j \mid r_j = \arg\max_r (S(r, q_j)), r \in \mathbf{P}^j_{can}, \quad j = 1, 2, \cdots, n\}. \quad (5)$$

$\mathbf{P}^j_{\text{can}}$ is the set of candidate regions which has higher $CH(r_i, q_j)$ than threshold, θ.

$$\mathbf{P}^j_{\text{can}} = \{r \mid CH(r, q_j) > \theta\}, \quad (6)$$

where $CH(r_i, q_j)$ describes a similarity in color histogram between candidate region r and j th patch of query object q_j.

Second, $\mathbf{R}^i$ is defined as the set of candidate regions which satisfies geometric relation based on each element in $\mathbf{R}^{\max}$. $\mathbf{R}^i$ can be obtained by

$$\mathbf{R}^i = \{r_j \mid r_j = \arg\max_r (S(r, q_j) G(r, i, j)), \quad r \in \mathbf{P}^j_{can}, r_i \in \mathbf{R}^{\max}, j = 1, 2, \cdots, n, j \neq i\}. \quad (7)$$

In Eq.(7), $G(x, i, j)$ is given as

$$G(x, i, j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(position(x) - \mu)^2}{2\sigma^2}}, \quad (8)$$

Figure 4: Results of object recognition. The matched candidate regions in input images are outlined

where

$$\mu = position(q_j) - position(q_i) + position(r_i). \quad (9)$$

The mean, μ represents the location expected to be matched with q_j. By multiplying Gaussian function $G(x, i, j)$ to $S(r_k, q_j)$, we can obtain a similarity with geometric relation between r_k and q_i.

Finally, $\mathbf{R}^*$ is chosen in $\mathbf{R}^i$ in such a way that matching score is maximized as

$$\mathbf{R}^* = \arg\max_{\mathbf{R}^i} (V(\mathbf{R}^i)). \quad (10)$$

Here, $V(\mathbf{R}^i)$ is the matching score of $\mathbf{R}^i$. $V(\mathbf{R}^i)$ is obtained by

$$V(\mathbf{R}^i) = S(r_i, q_i) + S(r_i, q_i) \sum_{j=1, j\neq i}^{n} (S(r_j, q_j) G(r_j, i, j)). \quad (11)$$

Consequently, we can find the location of the query object in a input image through $\mathbf{R}^*$.

6 Experiment

We utilize 10 objects for object recognition. First, we establish object models for each object. We adopt front view point only. We took several images according to the changes of illumination, occlusion and the size of objects. The total number of test images was 600.

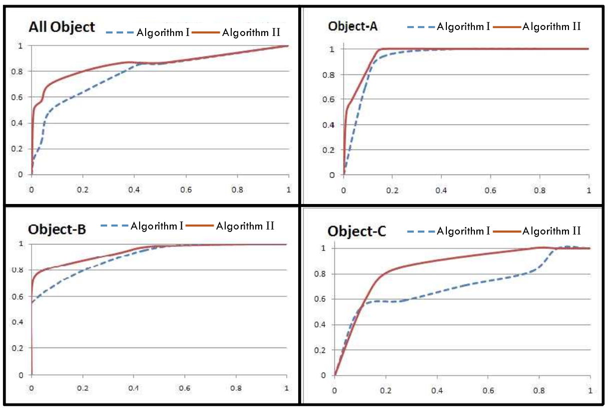

Figure 5: The ROC curves for objects

Fig. 4 shows the results of our matching process. The upper two images show the results for non occulding cases. The lower two images show the results for occlusion cases. Fig. 5 shows ROC curves for objects. Spatial relation is not applied in Algorithm I. On the other hand, it is applied in Algorithm II. It is observed from Fig. 5 that Algorithm II is better than Algorithm I. For the occlusion case, 83 objects out of 120 are recognized.

7 Conclusion

We proposed the method for object recognition. In our method, object models are divided into many patches. And, Color CH are used as a feature for each patch. In addition, we showed that a better performance of object recognition can be obtained by using the spatial relationships among patches. For one of our future works, we are going to develop a rotation-invariant method and a more efficient method in such a way that less number of representative colors are employed to reduce computational burden.

Acknowledgements

This work was performed by the Seoul R&BD Program and a grant of the Second Stage of BK21 Research Division for Advanced IT Education Program on Industrial Demand.

References

[1] P. Chang, J. Krumm(1999), "Object Recognition with Color Cooccurrence Histograms," *IEEE Conference on CVPR 1999*, pp.2498, Fort Collins, CO, June 23-25, 1999

[2] S. Redfild, M. Nechyba, J. G. Harris and A. A. Arroyo (2001), "Efficient Object Recognition Using Color," *2001 Florida Conference on Recent Advances in Robitics*, Boca Raton, May, 2001

[3] A. Hegerath, T. Deselaers, and H. Ney (2006), "Patch-based object recognition using discriminatively trained Gaussian mixtures," *BMVC 2006*,pp.519-528

[4] D. J. Crandall and D. P. Huttenlocher (2006), "weakly supervised learning of part-based spatial models for visual object recognition," *LNCS 3951*, pp.16-29

[5] S. Thrun, W. Burgard, D. Fox (2005) , *Probabilistic Robotics*, pp85-116

[6] A. Ahmadyfard and J. Kittler (2000) , "region-based representation for object recognition by relaxation labeling," *LNCS* 1876,pp.297-307

[7] T. V. Pham and A. W.M. Smeulders (2006) , "learning spatial relations in object recognition," *pattern recognition letters 27* , pp 1673-1684, 2006

[8] X. Wang, J. M. Keller and P. Gader (1997) , "Using spatial relationship as features in object recognition," *NAFIPS 1997*, pp.160-165

[9] K. Ohba, Y. Sato, K. Ikeuku (2000) , "Appearance-based visual learning and object recognition with illumination invariance," *Machine vision and applications 2000*, pp. 189-196

[10] A. Wu, D. Xu, X. Yang, and J. Zheng (2005) , "Generic solution for image object recognition based on vision cognition theory," *Fuzzy Systems and Knowledge Discovery*, vol.3614, pp.1265-1275

[11] W. Zhang, H. Deng, T. G. Dietterich and E. N. Mortensen (2006) , "A hierarchical object recognition system based on multi-scale principal curvature regions," *ICPR 2006*, pp.778-782.

[12] G. Bouchard, B. Triggs (2005) , "Hierarchical part-based visual object categorization," *proceedings of the 2005 IEEE CVPR*, pp.710-715

[13] B. Epshtein and S. Ullman (2005), "Feature Hierarchies for object classification," *ICCV 2005*, pp.220-207