Artif Life Robotics (2009) 13:488–492
DOI 10.1007/s10015-008-0614-5

ORIGINAL ARTICLE

**Heebeom Bang · Sanghoon Lee · Dongjin Yu
Il Hong Suh**

Robust object recognition using a color co-occurrence histogram and the spatial relations of image patches

Received and accepted: July 10, 2008

Abstract We propose a robust object recognition system where patch-based pyramid images and the spatial relationships among patches are utilized for our image model. In particular, both a color histogram (CH) and a color co-occurrence histogram (CCH) are applied to obtain image features for each patch. The locations of subregions 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 the spatial relationships among patches. To show the validity of our proposed method, we employ input images from various environments as test images.

Key words Color histogram · Color co-occurrence histogram · Patch · Pyramid image · Spatial relationship

1 Introduction

An object recognition system has to be dependable, especially for a service robot. The object recognition is not an easy task, since there are changes in 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 in environmental conditions.

Many researchers have proposed novel models which would be robust to changes in environmental conditions. Chang and Krumm[1] proposed a model using a color co-occurrence histogram. This model contains not only a color histogram, but also geometric information. In addition, the model was built by images taken from various viewpoints to build a model which was invariant with rotation. Although their method was somewhat invariant to rotation, it was not reliable for the scale changes. The selection of a color model is crucial for color-based object recognition. Redfield et al.[2] utilized only 16 colors to describe an object, and implemented the matching process with those colors. They assumed that the illumination conditions were fixed. This would cause their method to be unreliable for changes in environmental conditions. Zhang et al.[3] proposed an object recognition system based on multiscale affine invariant image regions. Elsewhere, a part (patch)-based model has been applied to build a model robust to partial occlusions, where relationships among the patches are utilized to solve partial occlusions. It has been reported that the recognition performance could be improved by using those patch-based models.[4–8]

Some research work has focused on developing robust matching methods, instead of proposing novel models. This proposed learning-based object recognition methods for better performance.[9–11,13] Learning from various environments can improve performance, but it needs a lot of time to ensure reliable learning performances.

Here, we propose processes of both object modeling and object matching to be as robust as possible to changes in environmental conditions. Changes in illumination, rotation, and partial occlusions are examples of 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 a color histogram and a color co-occurrence histogram are used as features for each patch. We also employ representative colors in hue, saturation, value (HSV) color space. On the other hand, for matching, the locations of candidate regions which are similar to the reference image patches are found by a particle filter in all the input images. Our system recognizes an object by comparing candidate regions and patches. Moreover, spatial relationships among patches and candidate regions are also included for our matching process.

H. Bang · S. Lee · D. Yu · I.H. Suh (✉)
Department of Electrical and Computer Engineering, Hanyang University, Seoul 133-791, Korea
e-mail: mintand@pusan.ac.kr

This work was presented in part at the 13th International Symposium on Artificial Life and Robotics, Oita, Japan, January 31–February 2, 2008

To show the validity of our proposed method, we input images from various environments as test images.

2 Overview of the proposed robust object recognition system

Object recognition methods can be classified as a top-down approach or a bottom-up approach. In a top-down approach, we find an unidentified object in the input image. In a bottom-up approach, we examine which 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 then find out whether the unidentified object is in the input image or not. Finally, we identify the highly probable location of regions similar to these models in the input image.

2.1 Process of object modeling

Figure 1 shows the process of building an object model. First, we take object images from several viewpoints. Then we transform these object images into an image pyramid. During quantization, we transform red, green, blue (RGB) colors into quantized colors in the HSV color space. We then divide these images into regions which have the same size. The regions we take from this step are called patches. Consequently, each scaled image is divided into a number of patches. For each patch, information about its viewpoint, scale level, position, color histogram (CH), and color co-occurrence histogram (CCH) are recorded.

In the modeling process, a single object is divided into two-octave pyramid images. The size of the original object image is 640×480 pixels. The original image is divided into nine patches which are all the same size. The next octave image, which is half the size of the original one, is divided into four patches. Consequently, a single object is described by 13 patches, and each patch is described by a CH and a CCH. For both CH and CCH, the color data is defined in the HSV space. In addition, we utilize 16 representative colors for both the CH and the CCH. Finally, the distance data in the CCH is quantized from 0 to 12.

2.2 Matching process

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

A block diagram of our matching process is shown in Fig. 2. First, we choose a specific unidentified model from our database. Then we select a candidate region in the input image. The candidate region is an area in the input image which has to be closely matched with a patch of an object model. As the 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 the locations of the candidate regions need to be estimated correctly. 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 the candidate regions are compared with the features of the unidentified object. Among the candidate regions, those which have similar feature values to the specific patches are selected. For the candidate regions selected, we examine the spatial relationships among their neighboring candidate regions. Consequently, the location of the unidentified object in the input image is finally determined by matching a feature value and by comparing the spatial relationships among neighboring candidate regions.

2.3 Particle filter

The particle filter is used to search for highly probable candidate regions where parts of the unidentified object can be found. Since we use the spatial relationships among candidate regions, we can get the highly probable candidate regions.

A particle filter[12] is a kind of Bayesian filter which computes the posterior distribution by using several weighted samples. The Bayesian filter consists of two steps: prediction and update. In the prediction step, a probabilistic system transition model is used in a given observation to predict the posterior at time *t*.

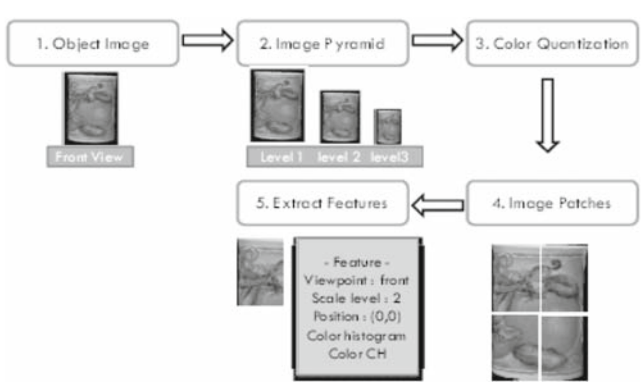

Fig. 1. Flow chart of object modeling process

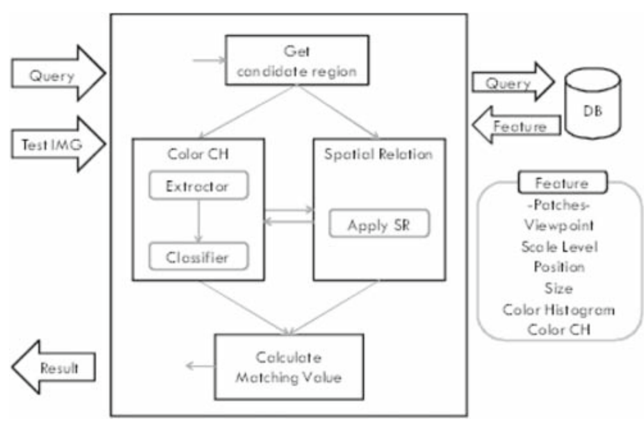

Fig. 2. Block diagram of matching process

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

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

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

In a particle filter, the posterior is approximated by a finite set of N samples $[\{x\}_t^i\}_I = 1, \ldots, N$ with importance weights ω_t^i. Sample x_t^i is drawn from the weight of the samples. The weights become the likelihood of observation.

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

To reduce the computational burden, the 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 size. If the sizes of the patches are different, the sizes of particles must be different in accordance with those of patches. In addition, the value of the color histogram from each particle is reusable.

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

3 Color co-occurrence histogram

3.1 Color model

There are various color models. Each color model has its own characteristics. The HSV color model consists of three components, hue, saturation, and intensity. The value of the hue is generally defined from 0° to 360°. The value of both the saturation and the intensity are defined from 0 to 1. The advantage of the 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 the RGB color space, can be defined as a representative color (C(red), C(black), etc.).

3.2 Color co-occurrence histogram

We use a color co-occurrence histogram as a 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 (Δx,Δy).We can write the color co-occurrence histogram symbolically as CCH (C1, C2,Δx,Δy). In order to make the CCHs invariant to rotation in the image plane, we ignore the direction of (Δx,Δy) and keep track only of the magnitude d given by

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

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, 15 representative colors are employed as the set of color C, and distance D is discretized as one of 12 integer values, as described by Chang and Krumm.[1]

3.3 Similarity

The similarity between a patch of the unidentified object and a candidate region in the input image can be computed as

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

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

4 Spatial relations

4.1 Modeling process

The object models are divided into patches at their different scale levels. Each patch has one independent feature. Consequently, we have a multifeature map. However, a single patch cannot describe a single object. A set of patches is required to describe a single object. Here, we consider the spatial relationships among patches by using their geometric relations.

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

Fig. 3. Five viewpoints of object model (Top, Left, Right, Front, Back)

By using the geometric relationships among those patches, the matching performance can be considerably enhanced.

4.2 Matching process

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

$$\boldsymbol{R}^{\max} = \left\{ r_j \middle| r_j = \arg\max_{r} \left(S(r, q_j)\right), r \in P^j_{can},\ j = 1, 2, \ldots, n \right\}. \quad (5)$$

where P^{can}_j is the set of candidate regions which have a higher $CH(r,q_j)$ than the threshold, θ.

$$P^j_{\text{can}} = \left\{ r \middle| CH(r, q_j) > \theta \right\}, \quad (6)$$

Here $CH(r,q_j)$ describes a similarity in the color histogram between candidate region r and the j-th patch of the unidentified object q_j.

Next, $\boldsymbol{R}^i$ is defined as the vector of candidate regions which satisfies the geometric relationship based on each element in $\boldsymbol{R}^{\max}$. $\boldsymbol{R}^i$ can be obtained by

$$\boldsymbol{R}^i = \left\{ r_j \middle| r_j = \arg\max_{r} \left(S(r, q_j) G(r, i, j)\right), \right. \\ \left. r \in P^j_{\text{can}}, r_i \in \boldsymbol{R}^{\max}, j = 1, 2, \cdots, n, j \neq i \right\}. \quad (7)$$

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

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

where

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

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

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

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

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

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

Consequently, we can find the location of the unidentified object in an input image through $\boldsymbol{R}^*$.

5 Experiments

We use ten objects for object recognition. First, we establish object models for each object. We adopt the front viewpoint only. We took several images in order to allow for changes in illumination, occlusion, and the size of objects. The total number of test images was 600.

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

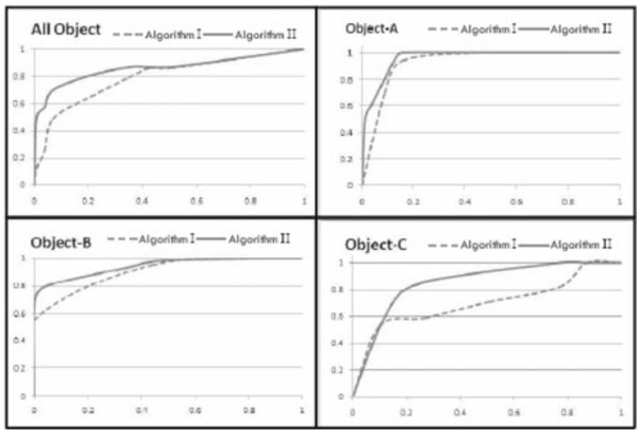

Fig. 5. The ROC curves for objects

Figure 4 shows the results of our matching process. The upper two images show the results for nonoccluding cases. The lower two images show the results for occlusion cases. Figure 5 shows the receiver operating characteristic (ROC) curves for objects. The spatial relation is not applied in algorithm I, but it is applied in algorithm II. It can be seen from Fig. 5 that algorithm II is better than algorithm I. For the occlusion case, 83 objects out of 120 were recognized.

6 Conclusion

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

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

References

1. Chang P, Krumm J (1999) Object recognition with color co-occurrence histograms. IEEE Conference on CVPR 1999, Fort Collins, CO, USA, June 23–25, p 2498
2. Redfield S, Nechyba M, Harris JG, et al (2001) Efficient object recognition using color. Florida Conference on Recent Advances in Robotics, Boca Raton
3. Zhang W, Deng H, Dietterich TG, et al (2006) A hierarchical object recognition system based on multi-scale principal curvature regions. In: Proceedings of the 18th International Conference of Pattern Recognition (ICPR'06) 2006, pp 778–782
4. Hegerath A, Deselaers T, Ney H (2006) Patch-based object recognition using discriminatively trained Gaussian mixtures. BMVC 2006, pp 519–528
5. Ahmadyfard A, Kittler J (2000) Region-based representation for object recognition by relaxation labeling. LNCS 1876, pp 297–307
6. Wang X, Keller JM, Gader P (1997) Using spatial relationship as features in object recognition. NAFIPS 1997, pp 160–165
7. Bouchard G, Triggs B (2005) Hierarchical part-based visual object categorization. Proceedings of the 2005 IEEE CVPR, pp 710–715
8. Epshtein B, Ullman S (2005) Feature hierarchies for object classification. ICCV 2005, pp 220–207
9. Crandall DJ, Huttenlocher DP (2006) Weakly supervised learning of part-based spatial models for visual object recognition. LNCS 3951, pp 16–29
10. Pham TV, Smeulders AWM (2006) Learning spatial relations in object recognition. Pattern Recognition Lett 27:1673–1684
11. Ohba K, Sato Y, Ikeuku K (2000) Appearance-based visual learning and object recognition with illumination invariance. Mach Vision Appl pp 189–196
12. Thrun S, Burgard W, Fox D (2005) Probabilistic robotics. The MIT, Cambridge, MA, pp 85–116
13. Wu A, Xu D, Yang X, et al (2005) Generic solution for image object recognition based on vision cognition theory. Fuzzy System Knowledge Discovery, vol 3614, pp 1265–1275