DYNAMIC ACTIVE SEARCH FOR QUICK OBJECT DETECTION WITH PAN-TILT-ZOOM CAMERA

Takahito KAWANISHI, Hiroshi MURASE, Shigeru TAKAGI and Martin WERNER *

NTT Communication Science Laboratories
3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa, Japan

ABSTRACT

This paper proposes a search method for detecting known objects quickly in 3D environments with a pan-tilt-zoom camera. In our previous work, we proposed an algorithm named Active Search that greatly reduces the number of calculations required to obtain a match between a reference object and an input image using color histograms. Here, we describe two improvements we have made to Active Search for such practical applications as robots and surveillance. First, we increased the robustness as regards the color histogram changes that result from different lighting conditions and camera angles by using multiple reference images and a pixel color vector quantization. Second, we reduced the number of camera operations (pan, tilt and zoom) by using a best-direction-first and upper bound pruning strategies. We call this camera control Dynamic Active Search. Experiments show an improvement in object detection accuracy and a 78% reduction in detection time.

1. INTRODUCTION

This paper proposes an algorithm for detecting target objects through pan-tilt-zoom cameras quickly and correctly. This algorithm is devised with practical applications such as robots and surveillance systems in mind. Conventional systems [1, 2] that use subtraction or template matching methods to detect objects are unsatisfactory in term of flexibility and performance. In 1996, we proposed a quick object detection method based on the color of a target object: "Active Search with color histograms"[3]. However, the method was insufficiently robust to handle the large changes in illumination conditions and camera angles that are unavoidable in practical applications. Moreover, the method was intended for use with static cameras without pan-tilt-zoom functions. Therefore, here we focus on improving robustness while maintaining quickness, and providing effective pan-tilt-zoom control mechanisms. To improve robustness, we introduce a pixel color vector quantization and multiple reference images taken under various conditions. To maintain quickness, we add an efficient algorithm to Active Search method to handle multiple reference histograms. In addition, we use a best-direction-first control strategy to reduce the number of pan-tilt-zoom operations.

*He is with Kugliga Tekniska Högskolan in Sweden. He contributed this work while studying at NTT

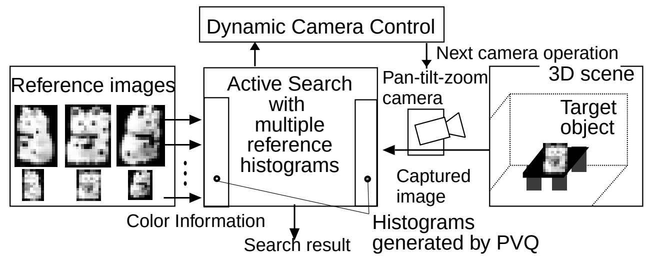


Fig. 1. Dynamic Active Search system

2. DYNAMIC ACTIVE SEARCH

Figure 1 outlines the Dynamic Active Search system. With this system, we correct multiple reference images under various illumination conditions, poses, zoom rates, and camera angles. We encode each pixel color by Pixel color Vector Quantization (PVQ) and construct histograms. We also encode input images from the camera by PVQ and search for the target object using the enhanced Active Search method with multiple reference histograms. The system uses this search result to direct the camera to the most probable area and controls the zoom operation to detect the target.

2.1. Pixel Color Vector Quantization

With the original Active Search approach, the color histograms are obtained by evenly sampling along each RGB axis and counting the number of times each discrete color occurs in the image. However, these histograms cannot withstand large changes in illumination conditions. The idea of using PVQ for histogram construction may be more promising if we use many sample reference images under various conditions. We construct a PVQ representation for pixel color as follows.

1. Each pixel's RGB color bit length is reduced from 8 bits to 6 bits by simply removing the two least significant bits.

2. Representative color vectors of all pixel colors in the reference images are calculated by an iterative vector splitting method used in the LBG algorithm [4]. The number of representative color vector is set so that the variation in the distances between each pixel color and its nearest representive color vector falls below an empirically determined value σ.

3. The nearest representative color vector is found for each color. If the color distance is smaller than a pre-defined threshold θ, then the color vector number is assigned to the color as code, otherwise a special code "0" is assigned, meaning that the color is not contained in the reference images.

2.2. Quick Image Search Algorithm

2.2.1. Active Search

To detect and locate a reference object in the input image, Active Search [3] calculates the similarity between a reference histogram and a histogram for a focus region (cropped sub area) of an input image by histogram intersection. The intersection $S(R,F)$ of two histograms, R and F, is defined as follows:

$$S(R,F) = \sum_{i}^{N} \frac{\min(R^i, F^i)}{|R|}, \text{ where } |R| = \sum_{i}^{N} R^i \quad (1)$$

i is the code of the representative color and N is the number of codes. If the similarity value exceeds the predefined threshold, we conclude that the object is located in that focus region. If the similarity value $S(R, F_a)$ of the reference histogram R and a focus region F_a is far below the threshold, the upper bound of similarity value $S(R, F_b)$ of R and another focus region F_b that overlaps F_a is estimated as follows:

$$|R| \cdot S(R, F_b) < |R| \cdot S(R, F_a) + |F_b - F_a| \quad (2)$$

where $|F_b - F_a|$ denotes the number of pixels in F_b but not in F_a. Active Search calculates the pruning search space by using Eq.2. That is, Active Search skips focus regions whose calculated upper bound is lower than the threshold.

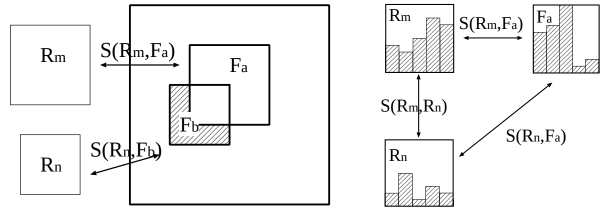
2.2.2. Active Search with Multiple Reference Histograms

A simple strategy for handling many reference histograms is to apply Active Search to each reference histogram. This paper proposes two, more efficient, algorithms : Active Search with Union Histograms and Parallel Active Search.

a) Active Search with Union Histograms

Union Histograms of the reference histograms are defined as:

$$U^i = \max\left(R_0{}^i, R_1{}^i, R_2{}^i, \cdots\right) \quad (3)$$



(a) Overlapping (b) Relationship between two histograms

Fig. 2. Parallel Active Search

Union Histograms have the following property.

$$S(U,F) \geq S(R_m, F) \quad (4)$$

Therefore, if Active Search is performed with these Union Histograms, a negative result (i.e. the target object is not found), guarantees the absence of the target object even if Active Search is performed with every reference histogram. In contrast, a positive result does not guarantee the presence of the target object. Therefore, candidate regions detected with Union Histograms are further checked with reference histograms.

b) Parallel Active Search

If two reference histograms R_m and R_n are given and the similarity value of R_m and R_n is not small, the similarity value $S(R_m, F_a)$ of R_m and a focus region F_a can be used to estimate the upper bound of the similarity value $S(R_n, F_b)$ of R_n and the other focus region F_b that overlaps F_a as follows:

$$\begin{aligned} |R_n| \cdot S(R_n, F_b) &\leq |R_m| \cdot S(R_m, F_a) \\ &+ |R_n| \cdot (1 - S(R_m, R_n)) \\ &+ |F_b - F_a| \end{aligned} \quad (5)$$

because $|R_n| \cdot S(R_n, F_a)$ is at most only the rest of R_n not including R_m larger than $|R_m| \cdot S(R_m, F_a)$ shown in Fig. 2(b). Therefore, the pruning area for the other reference histograms can be calculated during Active Search with a reference histogram R_m.

Reference images vary in size. The order of pixel counts of reference images ranges from 100 to $10,000$. The pruning area size also depends on the pixel count of the reference image. Usually, the bigger the pixel count is, the larger the pruning area becomes. Therefore, the system selects reference histograms in descending order of the pixel count. Recognition accuracy also depends on the pixel count. As the pixel count increases, the accuracy improves. We call histograms with a pixel count smaller than a threshold value c (around 900) "Candidate Detection Histograms" and histograms with a pixel count larger than c "Object Detection Histograms".

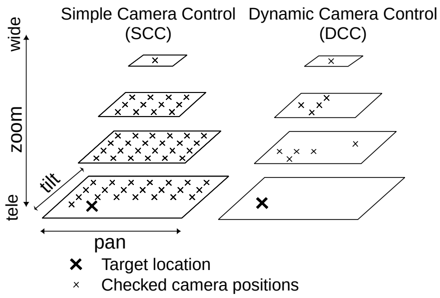


Fig. 3. Difference of camera control strategy

2.3. Dynamic Camera Control

Let the camera's maximum zoom depth be d, zoom steps be i, and zoom rate be δ. Then the total search space in an optimal pan-tilt-zoom position with no redundancy and sufficient capture becomes:

$$\text{total search space} = S\sum_{i=0}^{d}\delta^{2i} \qquad (6)$$

S is the search space at the widest angle. The camera's zoom rate δ is very small and d becomes very large. We select δ so that it is coarse enough to reduce the total search space and fine enough to maintain the object detection accuracy.

The total search space is 6 million pixels and more. To search effectively through this large space, we adopt the following search strategies: *(1) Wide angle first, (2) Best direction first.* The detailed algorithm is as follows:

Initialization:

1. Set the pan and tilt parameters to the top left corner.
2. Set the zoom parameter to the widest angle.
3. Perform the procedure: Search.

Procedure: Search:

1. Obtain an image. Apply Active Search with multiple reference histograms.
2. If one positive result is obtained by Object Detection Histograms, namely the target object is found, then finish.
3. If multiple positive results are obtained with Candidate Detection Histograms, sort the results by descending order of obtained focus region size and then similarity. For each obtained focus region result, perform a zoom operation and perform the procedure: Search.
4. If no match is found, select next pan-tilt position and perform the procedure: Search.
5. If all regions are checked and the target object is not found, then finish.

Fig. 4. Example reference images

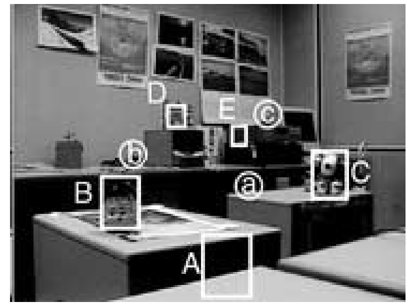


Fig. 5. Room environment

3. EXPERIMENTS

We conducted experiments to detect objects in a room environment. The specifications of the computer and camera we used are listed in Table 1. We select three objects for the experiments. For each object, we prepare about 100 reference images $(5$ different lighting conditions, 3 poses, $3{\sim}10$ zoom steps$)$. Example reference images for three objects are shown in Fig. 4. Figure 5 shows the room environment. Rectangles A, B, C, D, and E show the positions we used for the reference images. Circles a, b, and c show the positions we used for locating objects in our detection experiments.

3.1. Experiment 1: Search Accuracy

Object detection accuracy depends on both the pixel count of the objects and the pixel color coding scheme (PVQ and non-PVQ). Table 2 shows the experimental results. We used 15 images taken at the widest angle as test images. Here, the accuracy is the average of the precision rate and the recall rate. By using a reference image whose pixel count is greater than 900, the objects are perfectly detected with both PVQ and non-PVQ. The object detection accuracy with PVQ is more tolerant of a lower pixel count than that with non-PVQ. This implies that the number of useless candidate-directions detected with Candidate Detection Histograms can be reduced with PVQ.

Table 1. Experimental Specification

Computer	SGI O_2
CPU	R10000(250MHz)
Camera	Sony EVI-D30
Resolution	320×240

3.2. Experiment 2: Search Time

We measured the CPU time needed to search an input image for an object with multiple reference histograms (Table 3) for the following configurations:

1. non-PVQ + Active Search
2. PVQ + Active Search
3. PVQ + Parallel Active Search
4. PVQ+ Active Search with Union Histograms

There are two reasons for PVQ being faster than the original Active Search (non-PVQ) . First, PVQ reduces the number of histogram bins and there are fewer histogram intersection calculations. Second, PVQ is more discriminating than the original color scheme. The codes for PVQ tend to increase the similarity value within the reference histograms and the introduction of "0" code reduces the similarity value of a reference histogram and a focus region that does not contain the target. Thus PVQ reduces the upper bound similarity value of adjacent focus regions and increases the number of pruning regions. Moreover, there is a greater increase in the PVQ search speed for smaller reference images because the similarity value of the reference histograms decreases as the size of the reference images decreases and this decrease in similarity value with non-PVQ is more severe than that with PVQ.

In this experiment, the similarity values of the reference histograms are rather small and so the performance provided by Parallel Active Search is not large. Active Search with Union Histograms is very effective in reducing computation time, and the reduction in the search accuracy when using Union Histograms is not large.

3.3. Experiment 3: Total Search Time

We measured the total time needed to search the room environment for an object (Table 4) for the following configurations:

1. Active Search + Simple Camera Control(SCC)
2. Parallel Active Search + Dynamic Camera Control(DCC)
3. Active Search with Union Histograms + DCC

The time required for one mechanical camera operation was around 1.0 second. We performed about 100 camera operations with Simple Camera Control and half that number with Dynamic Camera Control. The number of camera operations with Union Histograms was nearly the same as that with Parallel Active Search. Most of the search time with Union Histograms was spent on camera operations. The detection accuracy was 100% for all configurations.

Table 2. Accuracy(average of precision rate and recall rate) of PVQ and non-PVQ

	$900 \sim$ pixels	$100 \sim 900$ pixels
non-PVQ	100%	57%
PVQ	100%	89%

4. CONCLUSION

In this paper, we proposed the Dynamic Active Search method, which can search for objects quickly in 3D environments. We improved the search efficiency by developing three extensions: Pixel color Vector Quantization, Active Search with multiple reference histograms and Dynamic Camera Control. We demonstrated our method for several objects in a room environment. The average search time was a fourth that with the original Active Search method and Simple Camera Control with no prediction. In the future, we will expand the Dynamic Active Search method from single pan-tilt-zoom camera search to multiple camera search.

5. ACKNOWLEDGEMENTS

The authors are grateful to Dr. K. Ishii, Dr. N. Sugamura, Dr. N. Hagita and Dr. E. Maeda of NTT Communication Science Laboratories for their help and encouragement.

6. REFERENCES

[1] T. Wada and T. Matsuyama, "Appearance sphere: Background model for pan-tilt-zoom camera," *Proc. of ICPR'96*, pp. 718–722, 1996.

[2] A. Rosenfeld and A. C. Kak, *Digital Picture Processing, 2nd ed.*, vol. 2, Academic Press, New York, 1982.

[3] V. V. Vinod and H. Murase, "Focused color intersection with efficient searching for object extraction," *Pattern Recognition*, vol. 30, no. 10, pp. 1787–1797, 1997.

[4] C. M. Huang and R. W. Harris, "A comparison of several vector quantization codebook generation approaches," *IEEE trans. on Image Processing*, vol. 2, no. 1, pp. 108–112, 1993.

Table 3. Search time for an image

	$900 \sim$ pixels	$100 \sim 900$ pixels
non-PVQ	2.7s	9.1s
PVQ	1.7s	3.9s
Parallel Search	1.5s	3.5s
Union Search	0.33s	0.35s

Table 4. Search time in room environment

	time
Active Search + SCC	$44.6s$
Parallel Active Search + DCC	$20.5s$
Union Histograms + DCC	$9.6s$