Systems and Computers in Japan, Vol. 31, No. 9, 2000
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J81-D-II, No. 9, September 1998, pp. 2035–2042

Fast Visual Search Using Focused Color Matching—Active Search

Hiroshi Murase and V.V. Vinod

NTT Basic Research Laboratories, Atsugi, Japan 243-0198

SUMMARY

There exist methods of searching for an object whose position and size are not known in an image. To obtain an accurate result by such methods, the focused region of an input image has to be compared many times with its reference image by varying its position and size. This makes fast searching difficult.

This paper describes a method of fast active search for an object and its position in an image by using the color histogram, which is stable to variations of shape of the object. When the input image of an object is compared with its reference image, the upper limit of the similarity in the image region is calculated from the input image in a certain position using the algebraic characteristics of the color histogram. If the upper limit is smaller than the search value, no search in this region is necessary, so that the number of matching operations can be greatly reduced. This method involves no approximations and has a computation speed about 10 to 1000 times faster than conventional exhaustive search methods. Applications of the proposed method, such as tracking, retrieval, and counting, are also described. © 2000 Scripta Technica, Syst Comp Jpn, 31(9): 81–88, 2000

Key words: Visual retrieval; object recognition; color information; image retrieval.

1. Introduction

With widening of multimedia environments, a technology to find an object and its position in an image has been in demand. This technology has various applications, for example, retrieval of a person or an object from an image [1–3], tracking of a moving object in a motion image, and recognition of an object against a complex background. To extract an object from an image, it is necessary to match two images by using their features. Use of the distribution of colors as a feature has become more popular recently since this feature is more stable to variations of the shape of an object. Swain's method [4] using color histograms is so effective for image retrieval that it has been widely used. However, when an object is in a part of an image (not over the whole image), the method of matching the whole image cannot be applied and another method that uses a focused region of the image has to be applied. There have been methods using data in a focused region of an image, for example, the method using colors in subblocks of an image [1] and the method using a pair of typical colors in adjacent subblocks [5]. These methods have fast processing speed but in principle are relatively low in accuracy.

To improve the accuracy of detection of an object in a focused region of an image, methods that use many focused regions have been proposed recently [6, 7]. These methods require a great many matching operations between input images and their reference image by varying the position and size of a focused region, resulting in a long retrieval time. To improve the processing speed, several solutions have been proposed: a method using a coarse resolution [9], a method using specific bins of color histograms, and a "coarse-to-fine method" [10] (fine examination follows coarse examination). These methods sacrifice accuracy for processing speed, and they do not find the combined maximum value of position and size in the whole parameter space. If an original image has a fine mixture of colors, the first search with coarse resolution cannot correctly represent its color histogram, so that a likely image region is not detected. The best resolution for the first search depends on each input image, and there is no clear guideline for some images. If a coarse color pattern (e.g., the French national flag) is detected by a fixed-size detector, the colors of the whole image will be recognized wrongly.

The proposed method, called the "active search method," is based on the idea that matching of a focused region is carried out at various sizes and at all positions but with measures to reduce the number of matching operations significantly. This method uses no approximations, so that the same accuracy as that of the exhaustive search method can be theoretically obtained. In other words, this method carries out a search for the maximum value of the similarity in the whole parameter space. The method can be combined with a conventional high-speed method in which the threshold value is controlled. Unlike other high-speed methods, the proposed method does not require adjustment of parameters for each object in order to maintain a certain accuracy since no approximation is involved.

A focused-color matching method is effective for detecting the approximate position of an object in an image since it uses color information alone. However, this sometimes requires another process to determine the accurate position, depending on the purpose of the detection. In such a case, it is possible to use the active search method to detect the approximate position of an object in an image, after which its final position is detected by using a template matching method. This paper includes a fast and accurate detection method, which is a combination of the active search method and a template matching method [11] that uses the low-frequency component of the discrete cosine transform (DCT) coefficient.

The proposed method has high processing speed and can be applied, for example, to the tracking of an object or the determination of its position in an image, and to retrieval in an image database, as shown in experimental results in this paper.

2. Focused Color-Histogram Method

The focused color-histogram method detects the position of an object in an image by comparing the color histogram of a focused region around it with a given reference image [7]. To process an object having an arbitrary size, a combination of the focused color-histogram method and the multiple-resolution representation method is used in this paper. In other words, this is a method in which the color-histogram scale (devised by Aswan [4]) is applied to a focused region of an image.

Figure 1 shows an example of an input image and its reference image. The proposed method is explained by using an exercise in which an image similar to a given reference image is extracted from an input image.

2.1. Formation of color histogram of reference image

A reference image is extracted from an image containing an object to be found. It is assumed that the extracted region is a square for simplicity, although this can be an arbitrary shape. Generally, a color histogram is a three-dimensional histogram consisting of red, green, and blue (RGB) space axes, each of which is divided into Q parts.

Let the histogram of the reference image be M_i', where $i = 1, \ldots, I$ and $I = Q^3$. Because of the properties of a histogram, $\sum_i M_i$ is the number of pixels in the region. Then, a normalized histogram $M_i = M_i' / \sum_j M_j'$ is obtained, with $\sum_{i=1}^{I} M_i$ being 1.0. Although this example of the color space uses RGB, another color space can be used if necessary, for example, intensity, hue, saturation (IHS). It would be possible to extract an object stably against the variation of brightness of the whole image. The proposed method is effective in reducing the amount of computation in cases where a color histogram other than RGB is used.

2.2. Multiple-resolutional representation

Generally, the size and position of an object to be found in the input image are not known. To find its position in an image, a square region is extracted and its histogram is calculated. The whole input image is scanned for this square region. To find the rate of the object in the input

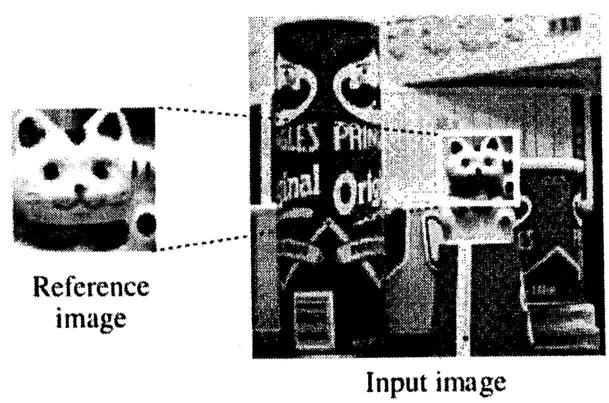


Fig. 1. A reference image and an input image.

image, the input image and the reference image are repeatedly compared by changing their relative sizes. For this operation, either the input image or the reference image can be changed. In the experiment in this paper, the former is changed by applying "multiresolutional representation." In this experiment, the size of the input image is changed with a step parameter of $\alpha = 1.25$. In practice, a parameter between 1.1 and 1.5 can be used for this experiment since the features of a histogram are not so sensitive to the size of the region. Figure 2 shows an example of multiresolutional representation.

2.3. Similarity (histogram intersection)

The "histogram intersection" proposed by Swain is used for evaluation of the degree of similarity, and for brevity is called "similarity" in this experiment. The similarity is given by
where H is the normalized histogram extracted from the

$$S_{HM} = \sum_{i=1}^{I} \min(H_i, M_i)$$

square region of the input image, and M is the normalized histogram of the reference image. S_{HM} is between 0 and 1 since it is normalized.

2.4. Detection of position of object in image

Assuming that an image contains a single object whose position and size are not known, let us find them by scanning the image with the said square region. The same operation is repeated for the input images having different resolutions. The position and size of the maximum similarity obtained from these operations are regarded as the final output values.

If there are multiple objects in an image, first the position of the maximum similarity is found, and a part corresponding to this region is eliminated from the image.

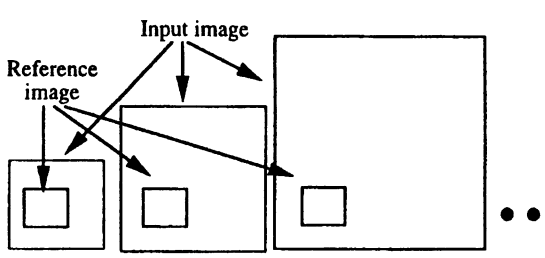


Fig. 2. An example of multiresolutional representation.

Then, the maximum similarity in the remaining region is found. This operation is repeated until the similarity value becomes less than some threshold value.

Focused-region color-histogram search methods are very effective for detecting objects in an image but require a very large amount of computation for histograms in various regions.

3. Active Search Method

The active search method can greatly reduce the number of searches compared with the exhaustive method having the same accuracy. In the active search method, a search is carried out by reducing the matching region relative to the upper limit derived from the histograms and similarity scale.

Generally, the similarity value of a focused region and reference image is often similar to that of its neighborhood. The coarse-to-fine method (fine scanning after coarse scanning) also uses this idea. However, if the collation between templates, the collation of many features, or the collation of distance value, is chosen as a feature, this assumption cannot be applied. For example, coarse scanning sometimes fails to perform the necessary detection because in a case such as a collation of templates, it is difficult to calculate the range of similarity in a neighborhood from the similarity in a certain position. Even if the range is calculated, it will not be narrow. However, if the features are histograms, and the similarity measure is intersection or a distance scale (similar to the former), the upper limit of the similarity can be calculated precisely and efficiently by algebraic methods. For example, the histograms of regions A and B in Fig. 3 differ only in the regions that do not overlap.

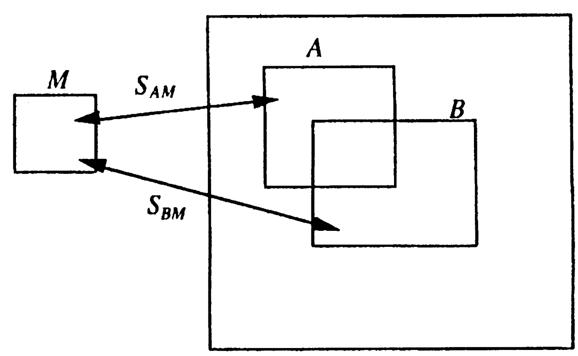


Fig. 3. Two intersecting focus regions in an image, and relation of histogram intersections for two regions.

3.1. Upper limit

Referring to Fig. 3, let regions A and B be arbitrary regions in an image, let M be a reference image, let S_{AM} be the similarity between A and M, let S_{BM} be the similarity between B and M, let $A \cap B$ be the common region of A and B, let A-B be a region where B is subtracted from A, and let $|A|$ be the number of pixels. When $|A| \leq |B|$, the following relationship holds:

$$S_{AM} \leq \frac{\min(S_{BM}|B|, |B \cap A|) + |A\text{-}B|}{|A|}$$

The right-hand term of this equation gives the upper limit of S_{AM}, when S_{BM} is calculated. If the similarity for a position is calculated, the upper limit of the similarity in its neighborhood is found. If this upper limit is less than the sought similarity, it is not necessary to calculate the similarity in this region. In this experiment, $|A| = |B|$ since the size of the region under consideration is not changed and the size of the input image is changed (see Section 2.2). However, the above equation for S_{AM} can be applied to its reversed case (the size of the region under consideration changes, and that of the input image does not change) since the equation generally holds for $|A| \leq |B|$. If there is more than one reference image (e.g., M and N), it is possible to use the relationship that the similarity between A and M is less than $\min(S_{AM}, S_{MN}) + \min(1 - S_{AM}, 1 - S_{MN})$. This can reduce the number of matching operations.

3.2. Active search using upper limit

Let us consider an algorithm to search for an object in an image using the above method, taking an example in which the size of an input image is changed and the similarity between the input image and the reference image is calculated at all positions.

Step 1: The size of the input image is converted to the same size as the reference image, and their similarity is calculated. This value is regarded as the initial value.

Step 2: The size of the input image is magnified α times (to keep a high resolution, in practice, the input image is reduced). If the size of the input image is greater than its maximum size (which is a parameter and is determined by the possible size of the object), go to Step 5.

Step 3: The similarity between the reference image and the region under consideration (the top left of the input image) is calculated.

Step 4: The upper limit is calculated so that search operations in a region where the target similarity exceeds the upper limit can be omitted. This means that the position of the object is moved into a region below the upper limit. If this similarity is greater than the target similarity, the former is replaced by the latter. Then the similarity around that region is calculated, and this operation is repeated. Go to Step 2.

Step 5: The position on which the similarity is maximum in each search position, and the size of the reference image at this instant, are regarded as the final results.

These result in a fine search around an object that is similar to the reference image, and a coarse search in the rest of the regions. Figure 4 shows an example of a search using the image shown in Fig. 1.

3.3. Reduction of computation

The proposed active search method aims to reduce greatly the amount of computation compared with the exhaustive method without reducing its resolution. Figure 5(a) shows the results of experimental comparison of three methods: the proposed method, an exhaustive method, and an approximate method. An INDY workstation (SGI) and the image shown in Fig. 1 were used for the experiment.

The coarse-to-fine method is used for the approximation method. First, a likely region was determined by scanning each direction of a three-dimensional space formed by the position and size parameters. The scanning is 4 times as coarse as in the exhaustive search method (density of 1/64 in the parameter space). Second, the likely region was scanned at fine intervals so that the final size and position could be detected. This method does not guarantee a correct result, and the extent of the coarse scanning is not logically determined (in practice, preliminary experiments are needed).

In the active search method, the search is carried out adaptively. Therefore, the amount of computation depends on each object and its background. Figure 5(b) shows the relationship between the number of searches and the area of the object, using the set shown in Fig. 1. When the

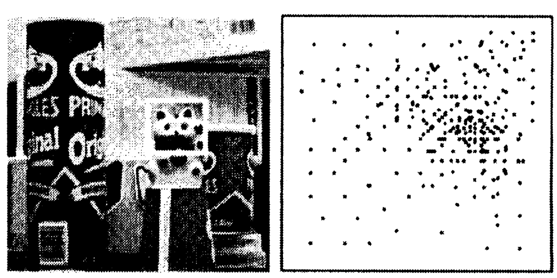


Fig. 4. An input image and searched points by the active search method.

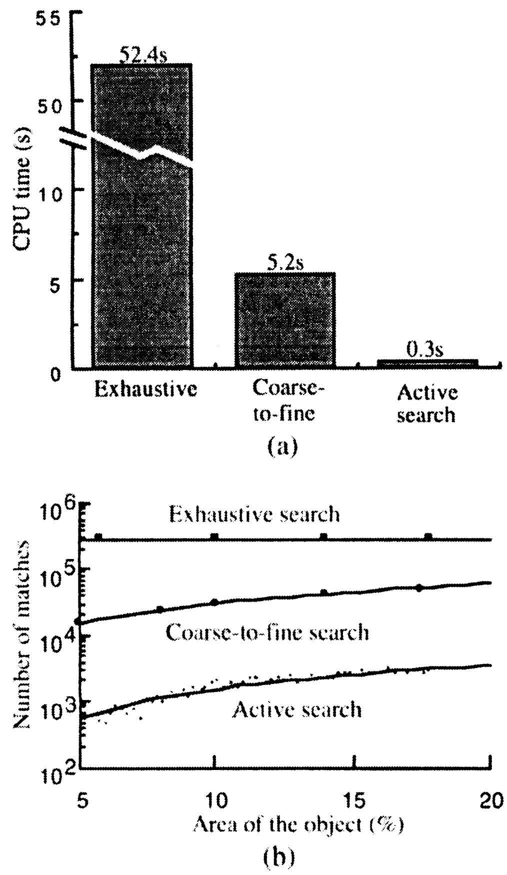


Fig. 5. Experimental results.

fraction of the image occupied by the object increases, the reduction factor of the computation time decreases.

4. Applications and Experiments

The fast image-search method has many applications. Applications of the active search method and their experimental results are described in this section.

4.1. Tracking of object

Tracking of an object (e.g., a human face) in a moving image using the active search method was investigated. The images were recorded at a rate of 10 frames per second using two sets of moving images.

- Data 1 (320 × 240 pixels per frame, 200 frames): A face was moved in front of a camera that moved at random.

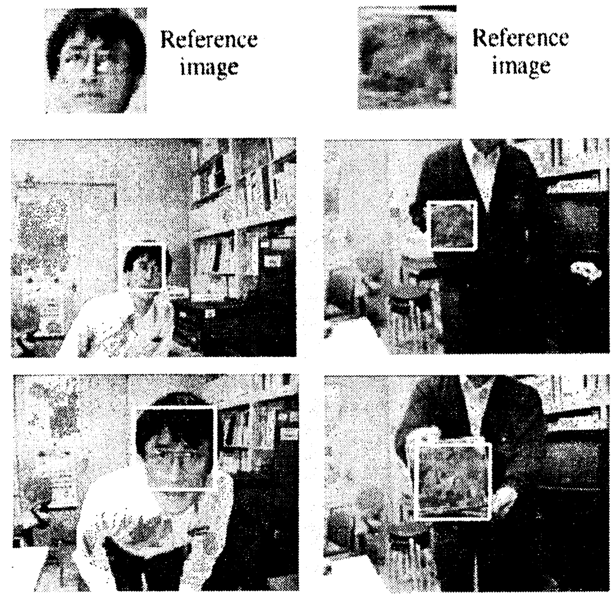


Fig. 6. An example of object tracking.

- Data 2 (320 × 240 pixels per frame, 200 frames): A paper package held in the hand was moved as shown in Fig. 6.

An INDY workstation (SGI) was used for the experiments. The position and size of the object were tracked at a mean rate of 64 ms per frame. Table 1 shows the efficiency and accuracy of the proposed method compared with the exhaustive search method.

4.2. Retrieval of image database

With a lowering of the prices of image input devices (e.g., Internet), abundant image data have recently become available, and fast retrieval of these data has become an important business. The proposed active search method performs such an activity.

Experiments searching for images similar to each reference image were carried out by providing three differ-

Table 1. Efficiency and accuracy of object tracking

	Time required for proposed method	Speed ratio versus exhaustive method	Error in position determination
Data 1 & 2	64 ms	103.7 times	4.7 pixels

ent databases, and they were compared with the conventional exhaustive search method. The test images are not shown in this paper due to copyright restrictions.

- Data 3 (240 × 240 pixels per frame, 60 frames): 12 indoor scenes.
- Data 4 (320 × 240 pixels per frame, 100 frames): 2 teddy bears.
- Data 5 (320 × 240 pixels per frame, 150 frames): Animated image ("Main character").

Table 2 shows the efficiency and accuracy of image retrieval by the proposed method, compared with the conventional exhaustive search method. "Precision" is the ratio of the correct output to the total amount of data, and "Recall" is the ratio of the correct output to the number of data that should be correctly retrieved. These are measures generally used for signal detection methods.

4.3. Counting of objects

Repeated retrievals are used for counting of objects.

- Data 6 (860 × 624 pixels per frame): Picture of fish (5 fish).
- Data 7 (320 × 240 pixels per frame, 600 frames): Basketball players wearing two kinds of uniforms.

Table 3 shows the efficiency of object counting compared with the exhaustive method. Figure 7 shows examples of counting. Figure 8 shows the accuracy of counting 600 frames of Data 7.

5. Discussion

A color histogram matching method like the proposed method uses the distribution of colors as its features. This is robust to noise and change of shape, such as occlusions, but possibly has a problem when a fine difference in the shape of an object is important. To solve this problem, a shape-matching method can be applied to a likely region that is preselected by the color histogram method. It is possible to detect the position of an object swiftly and accurately by applying a template matching method to a likely region that is preselected by a local-color matching method. In this experiment, a local DCT matching method was used for the second stage. This is a template matching method that matches the coefficients of low-frequency components, so that it is stable to high-frequency noise.

The relationship between the accuracy of detected position of an object and the amount of computation is

Table 2. Efficiency and accuracy of image retrieval

	Time required for proposed method	Speed ratio versus exhaustive method	Precision	Recall
Data 3	140 ms	15.5 times	1.0	1.0
Data 4	334 ms	17.19	0.8	1.0
Data 5	283 ms	34.18	0.79	0.85

Table 3. Efficiency of object counting

	Time required for proposed method	Speed ratio versus exhaustive method
Data 6	2002 ms	30.5 times
Data 7	581 ms	64.7 times

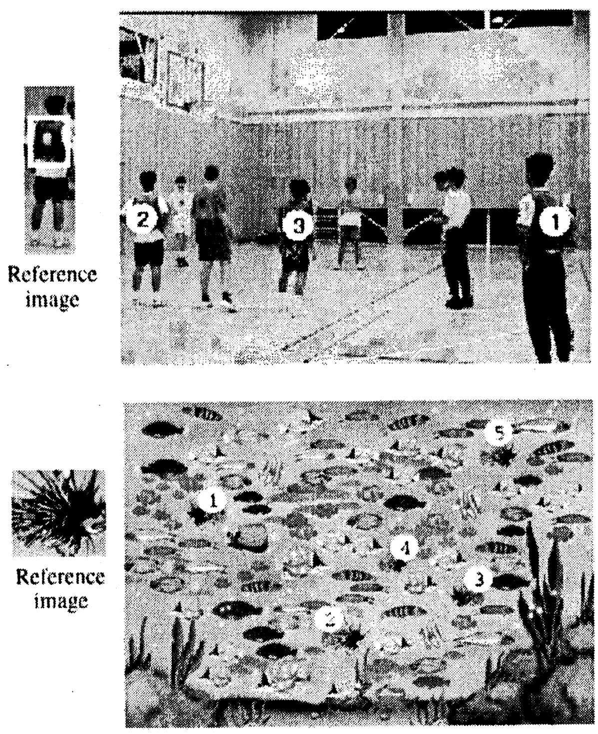


Fig. 7. Results of counting.

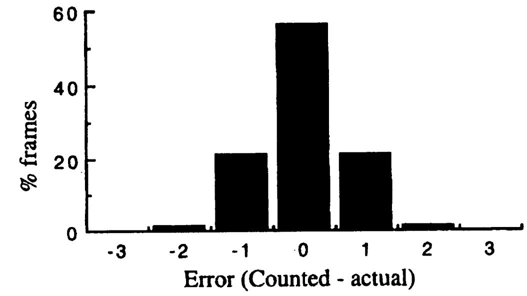


Fig. 8. Precision of counting.

experimentally compared by using Data 3 and four methods:

- Focused color DCT method (every 4 pixels and every 2 pixels)
- Focused histogram method (every 4 pixels and every 2 pixels)
- Back projection method (see Ref. 12)
- Proposed method (active search of every pixel, then local DCT search)

Figure 9 shows the results. A method like the local DCT method that uses a single matching between images has high resolution, but very fine scanning is required. If the scanning density is reduced, the accuracy of position determination becomes lower than in the focused histogram method. The proposed method, which is the more advantageous of the two methods, can achieve high accuracy and high speed.

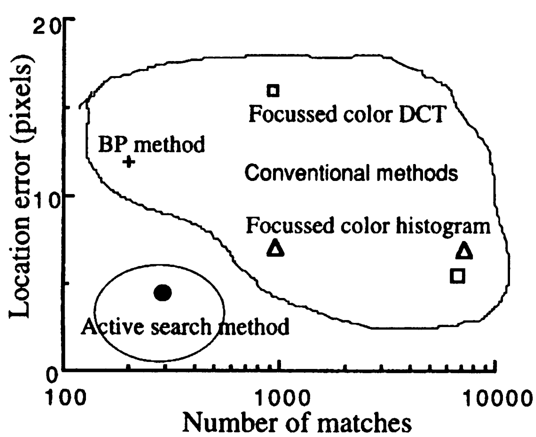


Fig. 9. Accuracy and computation for each method.

6. Conclusions

This paper proposes an "active search method" that searches for an object in an image at high speed. This method uses the algebraic properties of color histograms as a feature that is stable to variations of shape of the object. The search is carried out within an upper limit of the similarity of a position in the input image. This greatly reduces the number of matching operations, without using approximation. The experiments show that the computation speed of the proposed method is 10 to 1000 times faster than the exhaustive method.

The contents of this paper were presented at a meeting of the Pattern Recognition and Media Understanding Society in May 1997.

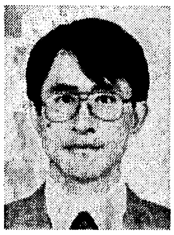
Acknowledgments. The authors express their thanks to Dr. Yoichi Higashikura (Director, NTT Basic Research Laboratories) and Dr. Ken-ichiro Ishii (Chief of Science Department, NTT BRL) for their support of this project and to Ms. Chie Hazume (NTT) for discussing this paper.

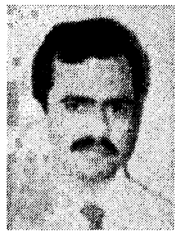
REFERENCES

1. Tamura H, Ikeda T. Intelligent information media. 1995.
2. Sakauchi M. Database vision and image retrieval. IEEE Multimedia, p 79–81, Spring 1994.
3. Niblack W, Barber R, Equitz W, Flickner M, Glasman E, Petkovic D, Yanker P, Faloutsos C. The QBIC project: Querying images by content using color, texture, and shape. SPIE, p 173–187, 1993.
4. Swain MJ, Ballard DH. Color indexing. IJCV 1991;7:11–32.
5. Nagasaka K, Tanaka Y. Automatic indexing and object search in color video images. 1992;33:543–550.
6. Murase H, Nayar S. Appearance-based detection of 3D objects in cluttered scenes. Pattern Recognition Lett 1997;18:375–384.
7. Vinod VV, Murase H, Hashizume C. Focused color intersection with efficient searching for image detection and retrieval. IEEE ICMCS 96, p 229–233.
8. Slater D, Healey G. The illumination-invariant recognition of 3D objects using local color invariants. IEEE Trans PAMI 1996;18:206–210.
9. Tomitaka C, Sekiya Y, Kageyama K. Object detection system for natural moving images. J Graphics CAD Soc 1993;63:13–19.

10. Ennesser F, Medioni G. Finding Waldo, or focus of attention using local color information. IEEE Trans PAMI 1995;17:805–809.
11. Vinod VV, Murase H. Object location using complementary color features: Histogram and DCT. ICPR96, p 554–559.
12. Swain MJ, Ballard DH. Indexing via color histograms. DARPA Image Understanding Workshop, p 623–630, 1990.

AUTHORS (from left to right)





Hiroshi Murase graduated from the Electronics Department of Nagoya University in 1978 and from the doctoral program in 1980. He then joined NTT. He has engaged in research on character/figure recognition, computer vision, and multimedia recognition. He was a guest researcher at Columbia University from 1992 to 1993. Presently, he is group leader of the Media Information Recognition Group (NTT Basic Research Laboratories). He holds a D.Eng. degree. He was awarded an Encouragement Award from IEICE in 1985, a Telecom System Technology Award from the Electrocommunications Promotion Society in 1992, a Best Paper Award from IEEE-CVPR in 1998, a Yamashita Memorial Award from the Information Processing Society in 1995, and the Best Video Award of IEEE-ICRA in 1996. He is a member of IEEE and the Information Processing Society.

V.V. Vinod obtained his Ph.D. degree from the Indian Institute of Technology in 1994. He was a Researcher at NTT Basic Research Laboratories from 1995 to 1997. He has been a Researcher at Singapore National University since 1997. He has engaged in research on neural nets and multimedia recognition.