

8—4 Object Search Using Orientation Code Matching

Farhan Ullah, Shun'ichi Kaneko and Satoru Igarashi *
Graduate School of Engineering
Hokkaido University

Abstract

A new method for object search is proposed. The proposed scheme is based on matching gradient information around each pixel, computed in the form of orientation codes, rather than the gray levels directly and is robust against irregularities occurring in the real world scenes. A probabilistic model for robust matching is given and verified by real image data. Experimental results for real world scenes demonstrate the effectiveness of the proposed method for object search in the presence of different potential causes of mismatches.

1 Introduction

Template matching is considered as one of the most powerful ways to search objects, for which we have many algorithms, programs and real systems for many applications. Many theoretical and practical approaches have been proposed such as template matching using half-tone images directly as a model [1, 2, 3], feature based approaches like Hough transforms which utilize edge pixels or lines as basic feature [4, 5, 6], or appearance based approaches like stochastic subspace matching using covariance information [7].

Conventional template matching techniques like correlation coefficient (CC) and sum of squared difference (SSD) are largely dependent on the brightness of the model and object image pixels [8]. CC is useful for avoiding the mismatch due to variations of illumination over the whole image, but partial occlusion, partial shading, background variations, target deformation or any combination of these irregularities cause misregistration. Matching using only the gradient magnitude can also give good results in cases of such irregularities if the brightness patterns differ by a constant factor, but the non-linear nature of the brightness variation can cause the gradient magnitude to vary as well, especially in cases of heavy shading or too bright highlighting. Gradient orientation information is by far the most invariant

feature in such cases as the difference of only a few gray levels among the neighbouring pixels can retain the orientation information.

Application of gradient information has been noted previously in tasks like finding dominant orientation [9] and gesture recognition [10]. Gradient information along with a relative distance information has been demonstrated to have good results in rotation invariant object recognition by constituting orientation tokens which are then used in the form of co-occurrence matrices [11]. In a recent paper it has been shown that the matching can be made robust by using only difference or tendency of adjacent pixel brightness values rather than their actual values [12]. This can be a useful example to show that it is effective and feasible to utilize not only brightness itself but also the accompanying information for object search.

In this paper we propose a technique, called Orientation Code Matching (OCM), for searching half-tone images by utilizing gradient information in the form of orientation codes which are matched optimally for realizing robustness [13, 14]. Orientation codes are generated by quantizing the angle corresponding to the steepest ascent orientation at each pixel. These codes, rather than brightness values directly, are utilized to find the object of interest. The proposed scheme has the advantage of being robust to abnormal brightness variations like highlighting or shadowing and/or variations of background. The algorithm is simple and can be implemented efficiently in real time systems.

Section 2 formalizes OCM and analyzes robustness of a similarity measure. Experimental results for real images are given in Section 3. We conclude the paper with some comments in Section 4.

2 Orientation code matching

2.1 Orientation code

In the OCM scheme, orientation code representations for an object image from the scene and the template are constructed from the corresponding gray images such that each pixel represents an orienta-

* Address: Kita 13, Nishi 8, Kita-ku, Sapporo 060-8628 Japan. E-mail: farhan@mee.co.in.eng.hokudai.ac.jp

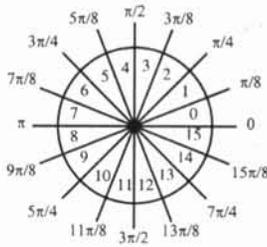


Figure 1: 16 Orientation codes ($\Delta\theta = \frac{\pi}{8}$).

tion code which is obtained by quantizing the orientation angle at the corresponding pixel position in the gray image. The orientation angle here refers to the angle indicating the steepest ascent orientation evaluated from the pixel neighborhoods, measured with respect to the horizontal axis. The orientation codes thus obtained are a function of the texture and shape of the object and hence essentially invariant to object translation and the effects of shading and background and illumination variations.

Suppose an analog image is represented by $I(x, y)$ and its horizontal and vertical derivatives as $\nabla I_x = \frac{\partial I}{\partial x}$ and $\nabla I_y = \frac{\partial I}{\partial y}$, respectively. For the discrete version of the image, they are evaluated around a pixel position (i, j) , and the orientation angle $\theta_{i,j}$ is computed as $\theta_{i,j} = \tan^{-1}(\nabla I_y / \nabla I_x)$. Since the numerical value of \tan^{-1} function is confined to $[-\frac{\pi}{2}, \frac{\pi}{2}]$, the actual orientation is determined after checking signs of the derivatives, thus making the range of θ to be $[0, 2\pi]$. The orientation code is obtained by quantizing $\theta_{i,j}$ into $N (= 2\pi/\Delta\theta)$ levels with a constant width $\Delta\theta$. Pixels with low contrast neighborhoods are sensitive to noise and hence designated by a separate code.

$$c_{i,j} = \begin{cases} \left\lfloor \frac{\theta_{i,j}}{\Delta\theta} \right\rfloor & : |\nabla I_x| + |\nabla I_y| > \Gamma \\ L & : \text{otherwise} \end{cases} \quad (1)$$

where Γ is a pre-specified threshold level for ignoring the low contrast pixels and L is a large value which is assigned as a code for them. An example of the orientation codes is depicted in Figure 1 corresponding to the quantization width of $\Delta\theta = \pi/8$.

We used the Sobel kernels for gradient angle computation because they are computationally efficient and are less sensitivity to additive noise because of the averaging of several pixels in the neighbourhood. In our setup we used 16 orientation codes corresponding to a resolution of $\frac{\pi}{8}$. The threshold level Γ was set to 10 as a tentative value but can be modified according to the nature of problem. L was set to 255 which is the maximum possible value for 8-bit representation of orientation codes.

2.2 Similarity measure

A similarity measure based on the definition of orientation codes is designed to evaluate difference between any two images of the same size. The best match between orientation code images of the template T and any object image I from the scene is searched by minimizing the summation of error functions as follows:

$$S_{m,n} = \frac{1}{M} \sum_{I_{m,n}} d(O_{I_{m,n}}(i, j), O_T(i, j)) \quad (2)$$

where $O_{I_{m,n}}$ and O_T are the orientation code images of the subimage at (m, n) and the template respectively, M is the size of the template and $d(\cdot)$ is the error function based on absolute difference criterion

$$d(a, b) = \begin{cases} \min\{|a - b|, N - |a - b|\} & : |a - b| < N \\ \frac{E_{max}}{2} & : \text{otherwise} \end{cases}$$

where E_{max} is the maximum possible error value between any two pixels.

When a comparison is performed between a pixel having an orientation code evaluated by the \tan^{-1} function and the one whose code was set to L due to low contrast neighbourhood, the error cannot be computed by finding the difference. In order to avoid such an inconsistent comparison, we need to assign a reasonable value to the error function corresponding to such pixels. The assigned value should be such that it does not bias the similarity evaluation for the subimage. For such cases, we assigned the value of $\frac{E_{max}}{2}$ to error function. This is the expected error value as explained later in 2.3. A large value for L is helpful for discriminating such an incompatible comparison.

Since the orientation codes are cyclic in nature, the absolute difference is not used directly for computing the error function, rather the minimum distance between the two codes is determined. As a consequence of this cyclic property of orientation codes, the maximum distance between any two codes is never more than $\frac{N}{2}$ which is assigned to E_{max} .

2.3 Robustness of similarity measure

In this section we provide a theoretical consideration of characteristics of the measure of similarity evaluation around any position where an object of interest is occluded by some other object. The analysis is based on a stochastic modeling of inconsistency between uncorrelated brightness of images, resulting in invariability of the measure proposed.

Suppose we have the fraction of occlusion represented by an occlusion rate β ($0 \leq \beta \leq 1$), then S can be separated into region of occlusion I_β and

non-occlusion $I_{1-\beta}$ in the target subimage as follows:

$$S = \frac{1}{M} \sum_I d(O_I(i, j), O_T(i, j)) \quad (3)$$

$$= \frac{1}{M} \sum_{I_{1-\beta}} d(O_I(i, j), O_T(i, j)) \quad (4)$$

$$+ \frac{1}{M} \sum_{I_\beta} d(O_I(i, j), O_T(i, j)) \\ = S_{1-\beta} + S_\beta \quad (5)$$

$S_{1-\beta}$ is the fraction of the total S evaluated over the region of non-occlusion, while S_β is the one evaluated over the occluded region. We omit the subscript (m, n) for simplicity. For $S_{1-\beta}$, we can expect a reasonable low value since the two partial images corresponding to the unoccluded region should be similar to each other. The main issue here is the analysis of S_β . It can be expressed as

$$S_\beta = \frac{M\beta}{M} \frac{1}{M\beta} \sum_{I_\beta} d(O_I(i, j), O_T(i, j)) \quad (6)$$

$$\approx \beta E_{I_\beta}(d(O_I(i, j), O_T(i, j))) \quad (7)$$

Here $E_{I_\beta}(d)$ refers to the expected value of d over I_β . The difference $d(O_I(i, j), O_T(i, j))$ has the supremum $N/2$ as mentioned earlier in 2.2. From a statistical reasonable assumption, in any occluded region, occluding objects have no relationship with the occluded objects and hence there is no correlation between $O_I(i, j)$ and $O_T(i, j)$. This makes the difference $d(O_I(i, j), O_T(i, j))$ to distribute uniformly, and the mean of this distribution is expected to be the constant $N/4$. Then the value S_β can be represented as $S_\beta = \beta N/4$. This invariability of S_β is very important to estimate the total value S for OCM. It is obtained as

$$S = S_{1-\beta} + \beta \frac{N}{4} \quad (8)$$

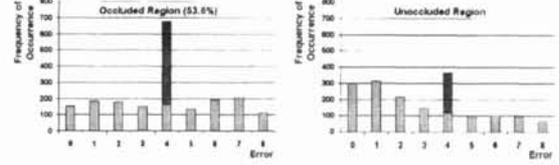
In general, according to this formalization, we can expect the same variation of S as that of $S_{1-\beta}$ which represents the property evaluated over any non-occluded part. This is the main reason that we can obtain much more robustness of OCM than some existing similarity measures. By using the above mentioned relation, an upper limit of occlusion rate β can be estimated as $\beta \leq \frac{4(T-S_{1-\beta})}{N}$ where T is a threshold value for verification of matching; for example, in the case that $N = 16$, $T = 3.5$ and $S_{1-\beta} = 1.5$, we can obtain the relation $\beta \leq 0.5$ which means any object image with an occluded region of up to 50% area of the whole size can be searched if similarity values for unmatched images are more than the threshold value $T = 3.5$.

Figure 2 shows example images adopted in verification experiments for our robustness analysis of



(a) Unoccluded Object (b) Occluded Subimage

Figure 2: Object and its occluded version.



(a) Occluded Region (b) Unoccluded Region

Figure 3: Distribution of error values.

OCM. 2(a) and (b) show an object image and its occluded version. For our experiments $N = 16$ and $E_{max} = 8$. Figures 3(a) and (b) show the distribution of all the error values over the occluded and unoccluded regions respectively. Frequency of the error values corresponding to $\frac{N}{4}$ is highest in both the images. It is due to the contribution of error values assigned to the low contrast pixels. This contribution is shown by the darker bars on the same plots. The histogram profile of the computed error values for the occluded region, shown in Figure 3(a), can be observed to be similar to a uniform distribution which is introduced in the above formalization. On the other hand, Figure 3(b) shows the error distribution in the unoccluded (normal) region and the profile is different from the one of the occluded region shown in Figure 3(a) and has distinct peaks near 0 and 1 levels.

The image is of size 58×68 and consequently 3696 pixels without peripheral pixels, and it involves the occluded area of around 1980 pixels, which constitutes about 53.6% of the total subimage region. The total accumulated error for the occluded region is 7833, and the mean is 3.96 which is almost the same as the expected value $\frac{N}{4} = 4$.

There are other similarity measures which are not robust because the similarity values over any occluded regions can not be estimated easily. For example, CC is a product-based correlation measure which makes it difficult to predict the correlation values in the occluded regions. SSD, on the other hand, is expected to have similar characteristics to OCM, but the range of difference in brightness is much wider than the one of orientation codes.

Table 1: Performance comparison

(a) Experimental setup

Total no. of images	150
Image size	240 × 320
Template size	68 × 58
Image type	8-bit gray scale
Camera	Victor GR-DVL7 digital video
CPU	AMD-K6 400MHz
OS	Windows 2000

(b) Comparative results

Method	Successes	Success rate	comp. time
SSD	13	8.1%	27.2 sec
CC	51	34.0%	58.8 sec
CC(L)	114	76.0%	54.8 sec
OCM	140	93.3%	39.2 sec

3 Experimental results

Table 1 shows a summary of the setup and results for the experiments carried out to check the effectiveness of the proposed method on a set of images. For this setup, a toy “Winnie the Pooh” was selected as the object of interest whose template image was taken separately. Various test images were prepared in which the toy appears in different conditions of shading, differing background, occlusion and their different combinations. We compared the results obtained from OCM with CC, SSD and CC(L). CC(L) here is the correlation coefficient based matching applied to the corresponding edge features obtained by using the Laplacian operator. Success means the best match is the same as the ground truths. OCM had the maximum number of successful matches followed by CC(L). This demonstrates the effectiveness of matching using only the gradient information or edge features rather than the brightness directly, especially in presence of real world irregularities.

Robustness of different matching schemes under various situations was checked using the images having some disturbances and irregularities. Images from the set used for comparison in Table 1 as well as other images were used in these experiments¹. We categorize the experiments according to the nature of problem as below.

3.1 Highlighting

A small part of a CD jacket was taken as a template and the true position in the scene is under high-

¹ Images used in the experiments and the setup described earlier are available at:
<http://mee.coin.eng.hokudai.ac.jp/open-pub/farhan/images.htm>



Figure 4: Comparative results for highlighted image.



Figure 5: Comparative results for test image with background shift and shading condition.

lighting as shown in Figure 4(a). Results by using CC, SSD, CC(L) and OCM are marked with boxes around the best match position for each method. As shown, OCM and CC(L) were able to locate the true position whereas mismatching occurred in cases of CC and SSD. This experiment demonstrates the fundamental effectiveness of OCM for large perturbation in brightness caused by illumination change. CC(L) is also robust in this situation.

3.2 Shading and background shift

Figure 5 shows an example image in which the object appears in different background compared with the template image and at the same time under partial shading. Only OCM was able to detect the object despite the disturbance occurring around the object.

3.3 Occlusion

Figure 6 shows the matching results from the four methods in the situation where the object is partially occluded by some other object. OCM could locate the object while the true position of the object was ranked at 45064th, 2504th and 25th in order of best match by using SSD, CC and CC(L) respectively.

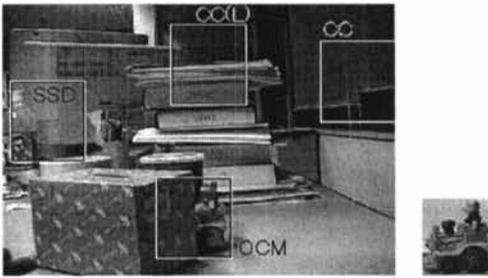


Figure 6: Comparative results for occluded object.

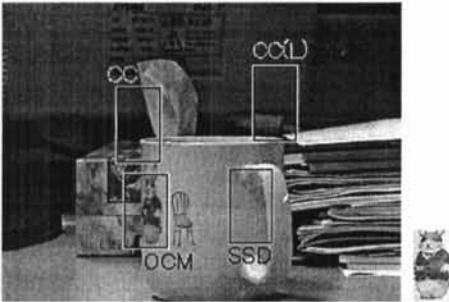


Figure 7: Comparative results for deformed object.

3.4 Deformation

Robustness against object deformation was checked by using an image of a cup shown in Figure 7 along with the template image. The object appears horizontally deformed due to rotation of the cup and forms only a part of the subimage. OCM could locate the object successfully.

4 Conclusions

Experimental results demonstrate the effectiveness of OCM for object search in real images even in the cases when some irregularity is present in the scene or when the object appears against different background. Results have also shown the robustness of OCM when the object only partially matches with the given template, which can be useful in extracting occluded objects or matching in cases of minor deformations. A probabilistic analysis has been provided for robust registration using OCM which has been verified by real image data.

OCM does not involve floating point computation during the matching operation and its computation cost is very low compared with the multiplication based methods like CC as can be seen from results in Table 1 which shows the computation time for the exhaustive search. OCM, being a minimization based search scheme, can be made even faster by

utilizing the sampling scheme proposed by [1] for faster registration.

References

- [1] D. I. Barnea and F. Silverman, "A Class of Algorithms for Fast Digital Image Registration," *IEEE Trans. on Computers*, vol. 21, no. 2, pp. 179-186, 1972.
- [2] W. K. Pratt, "Correlation Techniques of Image Registration," *IEEE Trans. on AES*, vol. 10, no. 3, pp. 353-358, 1973.
- [3] M. Svedlow, C. D. McGillem and P. E. Anuta, "Image Registration: Similarity Measure and Preprocessing Method Comparisons," *IEEE Trans. on AES*, vol. 14, no. 1, pp. 141-149, 1978.
- [4] D. H. Ballard, "Generalizing The Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111-122, 1981
- [5] D. P. Huttenlocher and S. Ullman, "Recognizing Solid Objects by Alignment with an Image," *IJCV*, vol. 5, no. 2, pp. 195-212, 1990.
- [6] F. Stein and G. Medioni, "Structural Indexing: Efficient 2-D Object Recognition," *IEEE Trans. on PAMI*, vol. 14, no. 12, pp. 1198-1204, 1992.
- [7] H. Murase and S. Nayar, "Appearance-based detection of 3D objects in cluttered scenes," *Pattern Recognition Letters*, vol. 18, pp. 375-384, 1997.
- [8] L. G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.
- [9] M. M. Gorkani and R. W. Picard, "Texture Orientation for Sorting Photos at a Glance," *Proc. 12th ICPR*, vol. 1, pp. 459-464, 1994.
- [10] W. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition," *IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition*, 1995.
- [11] V. A. Kovalev, M. Tetrou, and Y. S. Bondar, "Using Orientation Tokens for Object Recognition," *Pattern Recognition Letters*, vol. 19, no. 12, pp. 1125-1132, 1998.
- [12] I. Murase, S. Kaneko, and S. Igarashi, "Robust Matching by Increment Sign Correlation," *IEICE Journal*, vol. J83-D-II, no. 5, pp. 1323-1331, 2000 (In Japanese).
- [13] F. Ullah, S. Kaneko and S. Igarahi, "Object Search Using Orientation Histogram Intersection," *Proc. FCV2000*, pp. 110-115, 2000.
- [14] F. Ullah, S. Kaneko and S. Igarahi, "Object Recognition Using Orientation Code Matching," *PRMU*, vol. 100, no. 180, pp. 39-44, July 2000.
- [15] J. C. Russ, *The Image Processing Handbook*, CRC Press, 1995.