Hamming Distance Based Visual Tracking in Image Sequences

Prarinya Siritanawan[1], Toshiaki Kondo[1], Kanokvate Tungpimolrut[2] and Itsuo Kumazawa[3]

[1] Sirindhorn International Institute of Technology, Thammasat University, Thailand.

[2] National Electronics and Computer Technology Center, Thailand.

[3] Tokyo Institute of Technology, Japan.

Correspondence:
Prarinya Siritanawan
Sirindhorn International Institute of Technology, Thammasat University, Thailand.
Email: p.siritanawan@gmail.com

Abstract

The Hamming distance (HD) based gradient orientation pattern matching (GOPM) has been proven to be a robust object tracking technique against varying illuminations and occlusions in an image. The technique is baased on the template matching technique with unit gradient vectors (UGVs) instead of traditional image features such as intensities and gradients. This is due to the fact that UGVs are insensitive to illumination changes. Furthermore, to cope with the so-called occlusion problem, we search for a target pattern using the HD metric. The traditional HD metric, however, measures the similarity between patterns by the logical exclusive OR (XOR) operation and is suitable for binary encoded information only. This paper is concerned with the conversions of the similarities between UGVs to binary signals in order that the HD can be applied. We also investigate the effect of the threshold values used in the conversion.

Keywords: object tracking, gradient orientation pattern matching, Hamming distance, unit gradient vectors

1. Introduction

Typical visual tracking techniques often assume that the conditions of tracked objects are consistent, such that illuminations are constant and free from any outliers or occlusion [1]. Such assumptions are often violated in real situations, which cause erratic tracking. To cope with illumination changes, we previously devised the gradient orientation pattern matching technique (GOPM) [2][3]. Several implementations of the GOPM have proved its remarkable robustness to illumination changes and also its feasibility to perform at a video rate [4][5].

The GOPM is a template matching technique using unit gradient vectors (UGVs) as a robust feature. Instead of processing the gradient orientations with trigonometric functions like other gradient orientation based techniques, we make use of x and y components of the UGVs. The similarity measurement can be conducted with any traditional metric such as the sum-of-absolute differences (SAD), sum-of-squared differences (SSD), and zero-mean normalized cross correlation coefficients (ZNCC).

By introducing the Hamming distance (HD) as a new similarity metric, the technique can handle the occlusion up to 70% of the object. We call this newly devised technique the Hamming distance based GOPM technique [6].

In this research, we present different forms of applications of the Hamming distance to UGVs. We also study the effect of the thresholding parameters used in those applications.

2. Conventional Matching Methods

Conventional matching techniques based on image intensities measure the similarity between two patterns with various metrics, such as the sum-of-absolute differences (SAD), the sum-of-squared differences (SSD), and zero-mean normalized cross-correlation (ZNCC). The equations of those similarity metrics are expressed in Eqs. (1) to (3), where N is the size of the template, I^1 is the template (reference block), and I^2 is the second image where a best-matching block is being searched.

$$SAD(x,y) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} |I_1(x+i,y+j) - I_2(i,j)| \tag{1}$$

$$SSD(x,y) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} (I_1(x+i,y+j) - I_2(i,j))^2 \tag{2}$$

$$ZNCC(x,y) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I_1(x+i,y+j)-\bar{I_1})(I_2(i,j)-\bar{I_2})}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I_1(i,j)-\bar{I_1})^2}\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I_2(x+i,y+j)-\bar{I_2})^2}} \tag{3}$$

Let $d(x,y)$ be a distance of any matching methods at pixel coordinates (x,y). The best matching position of SAD or SSD is located at the minimum peak of any particular distance matrix as shown in Eq. (4), where A is the search area. In contrast to the ZNCC, the best matching position is given by the maximum peak of result distances.

$$[\hat{x},\hat{y}] = argmin_{x,y\in A} d(x,y) \tag{4}$$

3. Gradient Orientation Pattern Matching Method

The Gradient Orientation Pattern Matching (GOPM) is a template matching technique using unit gradient vectors (UGVs) as a robust feature. There are 2 main steps for performing the GOPM techniques; the first one is the feature extraction, UGVs, and the second one is the employment of matching method. The unit gradient vectors (UGVs) at the pixel coordinates (x,y) can be obtained through the following normalization (Eq.(5)):

$$n_x = \frac{I_x(x,y)}{\sqrt{I_x^2(x,y)+I_y^2(x,y)+\epsilon}} \qquad n_y = \frac{I_y(x,y)}{\sqrt{I_x^2(x,y)+I_y^2(x,y)+\epsilon}} \tag{5}$$

where I_x and I_y are the gradients in x and y directions. The gradients can be effectively approximated using Sobel operators. The ϵ is a small positive constant added to prevent zero-division during the normalization. The x and y components of the UGVs are visualized after scaling between 0 and 1.

Using the UGVs obtained above, the best matching position can be searched for with any traditional metric such as the sum of absolute differences (SAD), sum of squared differences (SSD), or zero-mean normalized cross-correlation (ZNCC). In our previous work, the similarity metric of gradient orientation pattern matching (GOPM) is based on either SAD or SSD metric [2]-[5]. The performances of these two metrics have similar results. For instance, the SSD for unit gradient vectors are expressed in Eqs. (6) and (7).

$$SSD_{n_x} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(n_{x1}(x+i,y+j) - n_{x2}(i,j))^2 \quad (6)$$

$$SSD\_n_y = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(n_{y1}(x+i,y+j) - n_{y2}(i,j))^2 \quad (7)$$

where SSD_n_x*(x, y)* and SSD_n_y*(x, y)* are the similarity indexes between the UGVs of a test image and the template at pixel coordinates *(x,y)*. Note that N is the size of template, n_{x1} and n_{y1} are the UGVs of the test image, and n_{x2} and n_{y2} are the UGVs of the template. The gradient orientation pattern matching (GOPM) is then given by Eq. (8).

$$GOPM(x, y) = SSD\_n_x(x, y) + SSD\_n_y(x, y) \quad (8)$$

GOPM (x,y) is equivalent to the squared Euclidean distance between two UGVs at the pixel coordinates *(x,y)*. Finally, the best matching position is located at the minimum peak of GOPM image as shown in Eq. (9) where A is the search area.

$$[\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}] = \boldsymbol{argmin}_{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{A}}\boldsymbol{GOPM}(\boldsymbol{x},\boldsymbol{y}) \quad (9)$$

Although the GOPM is robust to illumination changes, it cannot deal with the occlusion problem. This is due to the traditional similarity metric such as SAD and SSD. To cope with the occlusion problem, we have introduced the Hamming distance (HD) to the GOPM as a new similarity metric.

4. Hamming Distance Function

The Hamming distance (HD) is a similarity metric using logical exclusive OR operator (XOR). To measure the distance between 2 binary patterns, the HD compares 2 corresponding binary pixels and then take the summation over the pattern. General description of the HD metric is defined as Eq. (10).

$$HD(x,y) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} I_1(x+i,y+j) \oplus I_2(i,j) \quad (10)$$

where I_1 is the binary intensity of the test image at the pixel coordinate (x,y) and I_2 is that of the template. To implement the HD with non-binary information such as intensities, gradients, and UGVs, we need to apply thresholding to those non-binary measurements. For example, the HD may be used for intensities as [7] (Eq. 11).

$$sim(I_1, I_2, (x,y)) = \delta(|I_1(x,y) - I_2(x,y)| > \tau) \tag{11}$$

where τ is a certain threshold. When the absolute difference is greater than the threshold τ, we assign 1 to that pixel and 0 otherwise.

According to the ideas above, in this paper, we make use of UGVs, instead of intensities, with the HD metric. The diagram of tracking system is shown in Fig. 1. Thus, we can count the number of the UGVs that are not matched (Fig. 2). We have tested three different approaches for computing the HD as shown in sections 5.1. – 5.3.

4.1. Strategy 1: Hamming distance on UGVs with Arithmetic Combination

As illustrated in Fig. 2, the UGVs are converted to binary information and then calculated to the Hamming distance. Firstly, we calculate the distance results of x and y components by thresholding absolute differences as shown in Eqs. (12) and (13).

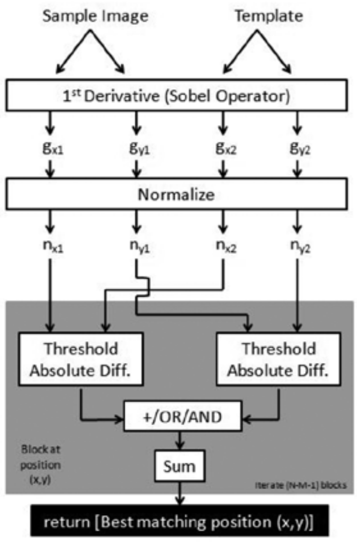


Fig. 1. Schematic diagram of the proposed method, gradient orientation pattern matching using the Hamming distance, with three variations; 1st strategy (arithmetic sum), 2nd strategy (AND operator), and 3rd strategy (OR operator).

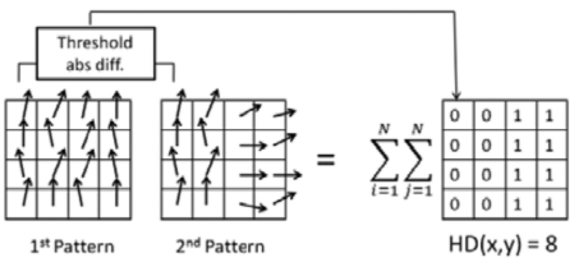


Fig. 2. Hamming Distance counts the number of UGVs that are not match.

$$d_x(i,j) = \begin{cases} 1, & |n_{x1}(x+i,y+j) - n_{x2}(i,j)| > \tau_1 \\ 0, & otherwise \end{cases} \quad (12)$$

$$d_y(i,j) = \begin{cases} 1, & |n_{y1}(x+i,y+j) - n_{y2}(i,j)| > \tau_2 \\ 0, & otherwise \end{cases} \quad (13)$$

where n_{x1} and n_{y1} are the UGVs of the input image, and n_{x2} and n_{y2} are the UGVs of the template, τ_1 and τ_2 are certain thresholds. The binary distance of UGVs in x and y components are then calculated by arithmetic summation in Eq. (14).

$$d(i,j) = d_x(i,j) + d_y(i,j) \quad (14)$$

Thus, the Hamming distance of the block at the coordinate *(x,y)* is then given by Eq. (15).

$$HDGOPM(x,y) = \sum_{i=1}^{N}\sum_{j=1}^{N} d(x+i, y+j) \quad (15)$$

This accumulative block above indicates the similarity between that block pattern and the template. Note that N is the same size as the template. This distance criterion is the same as the so-called city-block distance or D_4 distance. After obtaining the Hamming distance of UGVs, the best matching position is located at the minimum Hamming distance position as shown in Eq. (16).

$$[\hat{x}, \hat{y}] = argmin_{x,y \in A}\, HDGOPM \quad (16)$$

4.2 Strategy 2: Hamming Distance on UGVs with Logical AND Combination

The second strategy is another deviation of the first strategy. To combine the distance results of UGVs in *x* and *y* components, instead of arithmetic summation in Eq. (14), the threshold absolute difference function is calculated by using logical AND operator in Eq. (17).

$$d(i,j) = d_x(i,j) \wedge d_y(i,j) \quad (17)$$

As in Eqs. (15) to (16), the Hamming distance at the coordinate (*x, y*) are calculated and the best matching position is located at the minimal Hamming distance position.

4.3 Strategy 3: Hamming Distance on UGVs

The third strategy is an alternative approach of strategies 1 and 2. To combine the distance result of x and y components, the distance results are combined by OR operator as shown in Eq. (18).

$$d(i,j) = d_x(i,j) \wedge d_y(i,j) \quad (18)$$

The Hamming distance of the block at the coordinate (*x, y*) is then given by Eq. (15). The best matching position is also located at the minimal Hamming distance position as shown in Eq. (16). This distance criterion is the same as the so-called chessboard distance or D_8 distance.

5. Experimental Results

5.1 Matching Methods

A comparative study of matching techniques is conducted by calculating motion vectors in artificial image sequences under various conditions. Three artificial image sequences are created using three standard test images: Lena, Cameraman, and House. The size of the test images is of 512 by 512 pixels. The test image is used as the first frame in each image sequence, and the second frame is then generated by translating the first frame by 5 pixels in both vertical and horizontal directions. To compute motion vectors over the entire image, we partition the image into 31 by 31 (namely, 961) blocks of size 16 by 16 pixels. We allow 8-pixel marginal spaces at the top, bottom, left, and right parts of the image as the search area for matching (Fig. 3).

5.2 Illumination Change Test

We first evaluate the motion estimation performances of the three methods under irregular illuminations. The images are corrupted with 40dB Gaussian noise (Fig. 4(a)). The image intensities of the second frames are then non-uniformly changed as shown in Fig. 4(b). The intensities are reduced by half in both vertical and horizontal stripes. The intensities in the areas where two stripes overlap are reduced to one fourth.

Successful motion estimation rates (%) of 961 motion vectors estimated by the standard block matching techniques with the SSD metric (Fig. 4(c)), the GOPM with SSD metric (Fig. 4(d)), and the proposed method, the GOPM with the Hamming distance or HDGOPM (Fig. 4(e)) are shown in Table 1. It is obvious that the GOPM overcomes the lighting change problem (success rates approximately 99% by the SSD similarity metric and 100% success rates by the HD metric). Meanwhile, the traditional matching with the SSD metric is incapable of estimating image motions under irregular lighting (success rates are approximately 20%).

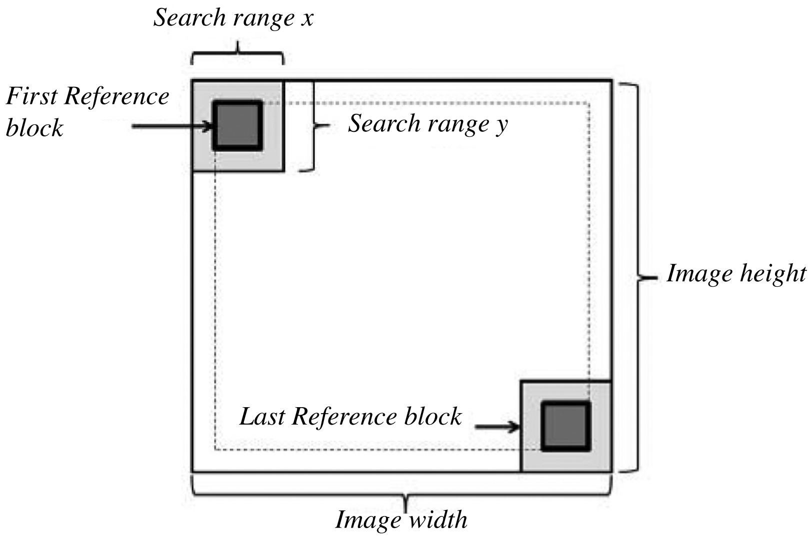


Fig. 3. 961 motion vectors are computed from the synthetic sequences. The reference block has size 16 by 16 pixels and the search range is defined on the area +/- 8 pixels around the reference block.

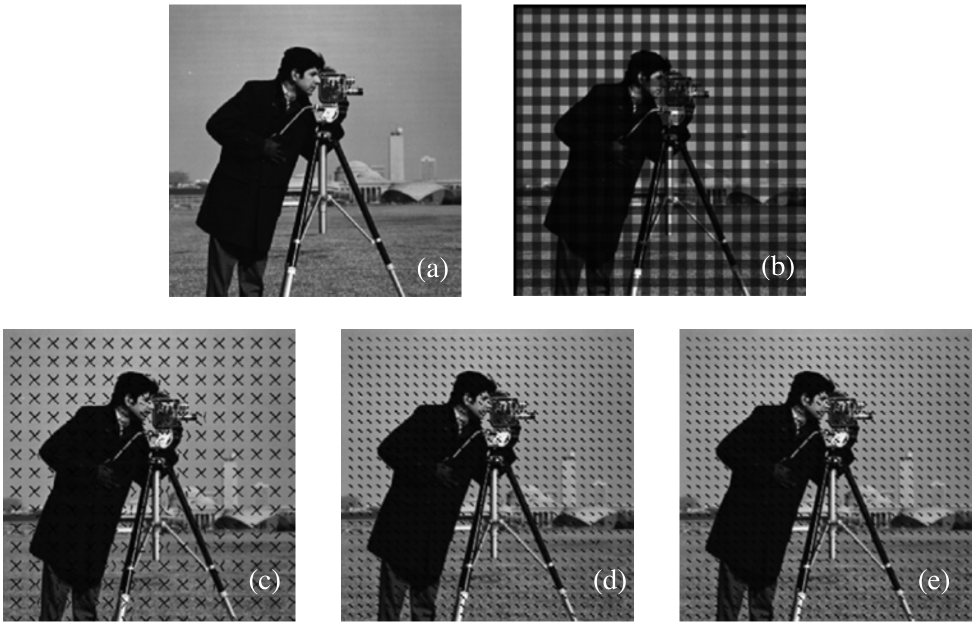


Fig. 4. Motion estimation results. (a) the first frame of a synthetic image sequence, (b) the second frame generated by translating the first frame by (5,5) pixels and subject to non-uniform changes of intensities, the motion vectors estimated by (c) the traditional block matching with the SSD metric, (d) GOPM with the SSD metric, and (e) GOPM with the Hamming distance (HDGOPM).

Table 1 Successful motion estimation rates (%) under non-uniform illumination change.

Image	*SSD*	*GOPM with SSD*	*GOPM with the HD*
Lena	21.23	99.89	100
Cameraman	26.74	99.89	100
House	23.10	99.37	100

5.3 Occlusion test

We next evaluate the motion estimation performances of the three methods under both constant lighting conditions with the so-called occlusion problem. Fig. 5 illustrates how part of a block in the second frame is covered with a black mask. The same occlusion is applied to 961 blocks equally. Fig. 6 shows the successful motion estimation rates under constant lighting when the size of the occlusion mask is varied from (0,0) to (16,16) pixels. The First (HDGOPM) yields the highest success rates, while the GOPM performs similarly well with slightly lower success rates. Meanwhile, the success rate of the traditional matching technique SSD is extremely low even if the amount of the occlusion is small. This indicates that the traditional SSD matching technique is severely sensitive to the occlusion problem.

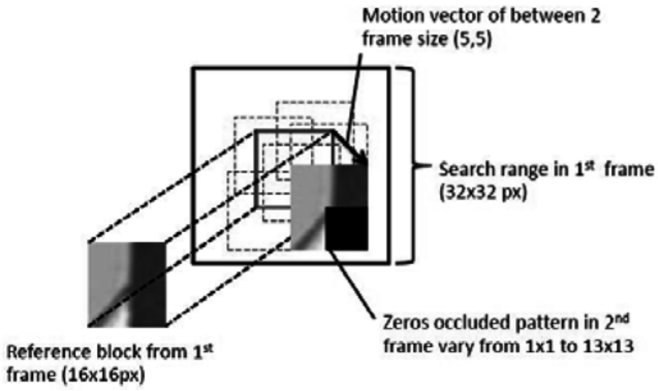
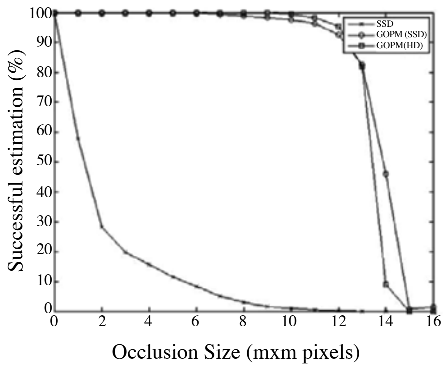
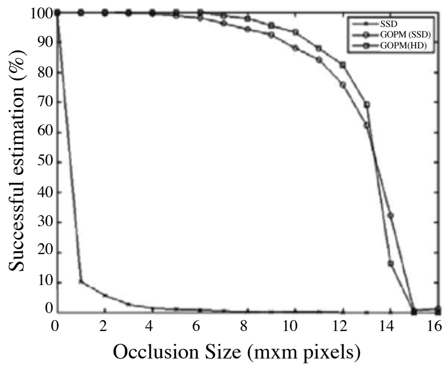


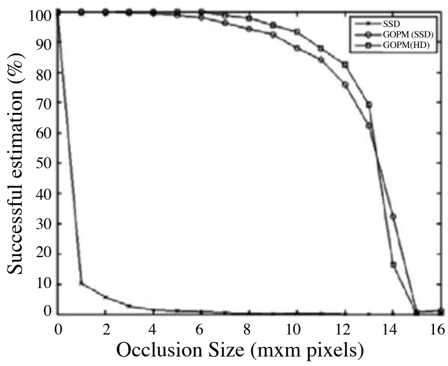
Fig. 5. Motion estimation on an artificial image sequence with (5,5) pixel translation, illumination change, 40dB Gaussian noise and occlusion with a dark mask.



(a) Lena



(b) Cameraman



(c) House

Fig. 6. Successful rate motion estimation of occlusion testing under uniform illumination on artificial motion sequences by using (a) Lena, (b) Cameraman, and (c) House image. The sizes of occlusion block are varied from 0×0 to 16×16 pixels.

5.4 Hamming Distance Representation

To test the tracking efficiency of each strategy, the successful motion estimation rates of the proposed method GOPM with the HD metric (HDGOPM) using Strategies 1, 2, and 3 (τ 1 and τ 2 are 0.3) under constant and varying illuminations are plotted against the amount of the occlusion (Figs. 7 and 8). It is apparent that Strategy 3 using a logical OR operation constantly outperforms the other two.

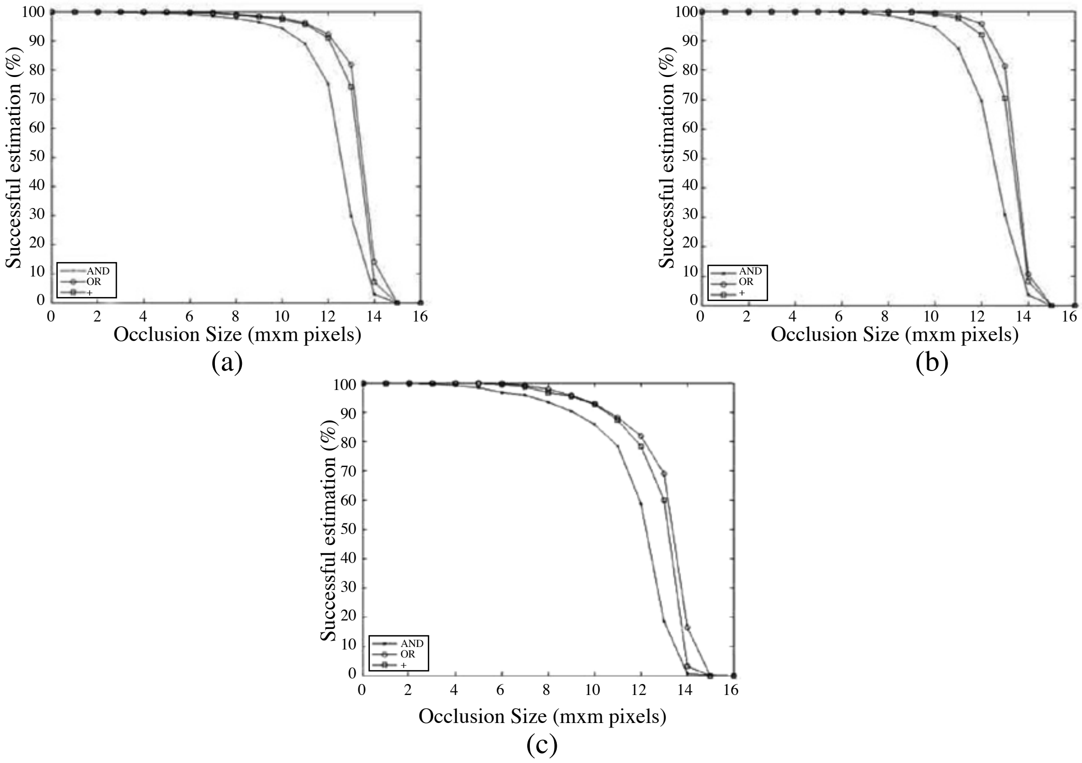


Fig. 7. Comparison of three strategies under constant lighting on (a) Lena, (b) Cameraman, and (c) House images. The vertical axis denotes the successful motion estimation rates among 961 blocks. The horizontal axis denotes the amount of occlusion from 0×0 to 16×16 pixels.

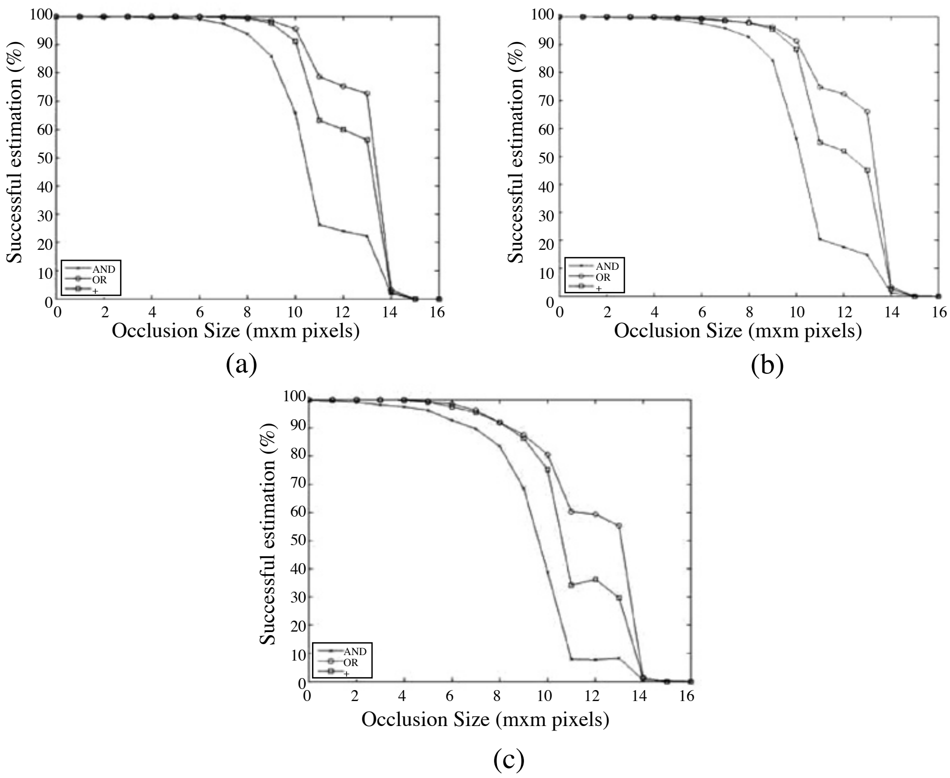


Fig. 8. Comparison of three strategies under non-uniformly varying lighting on (a) Lena, (b) Cameraman, and (c) House images. The vertical axis denotes the successful motion estimation rates among 961 blocks. The horizontal axis denotes the amount of occlusion from 0×0 to 16×16 pixels.

5.5 Threshold estimation

Prior to the use of the Hamming distance metric, thresholding is necessary for converting the similarities between two corresponding UGVs to binary codes. Here, we investigate the effect of the threshold value to the performance of motion estimation. For this, we again use the same three artificial image sequences and vary the threshold value from 0.1 to 2.0 with a step of 0.1. The motion estimation performance is tested under constant and varying illuminations with 40dB Gaussian noise. Figs. 9 and 10 shows that the successful motion estimation rates are fairly stable up to large threshold values, e.g., 1.6 for constant lighting and 0.8 for varying illuminations. Therefore, we can state that the performance is not heavily dependent on the threshold value. This means we do not need to select or adjust a threshold value, depending on input images.

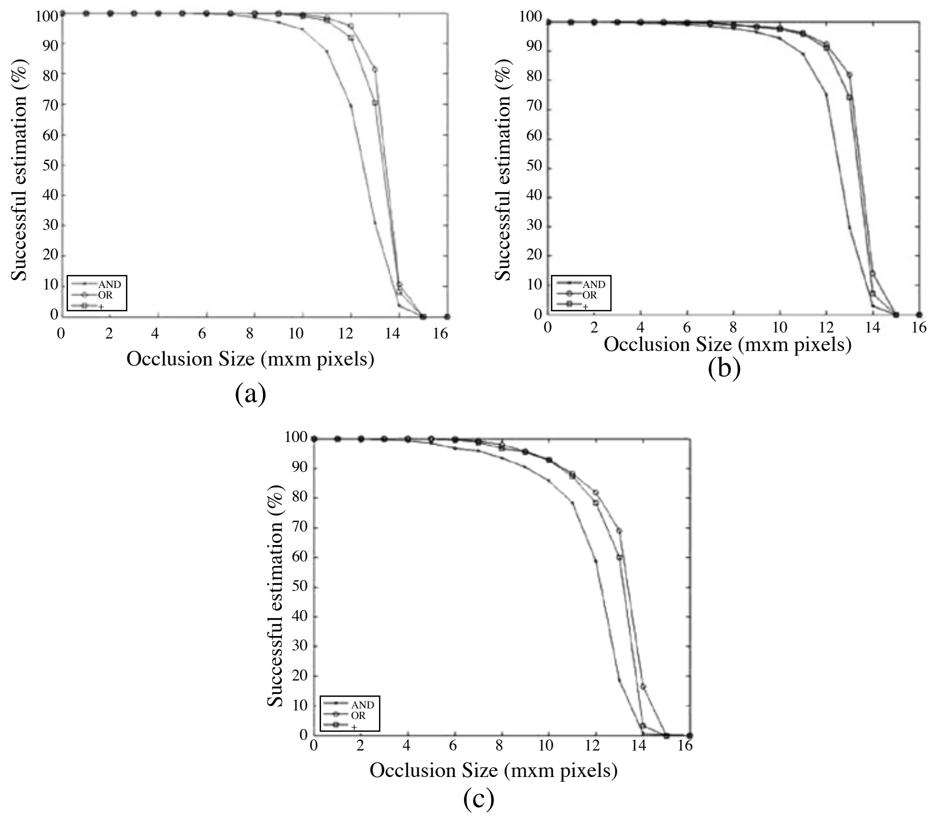


Fig. 9. Effect of the threshold value on motion estimation success rates under constant lighting on (a) Lena, (b) Cameraman, and (c) House images.

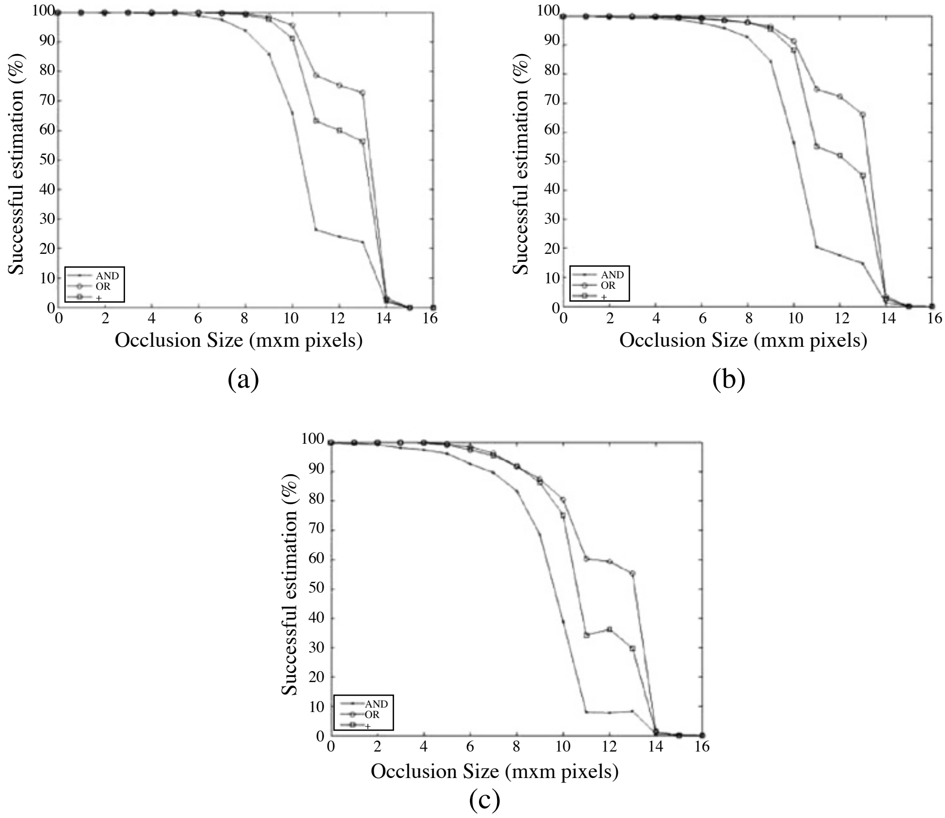


Fig. 10. Effect of the threshold value on motion estimation success rates under non-uniformly varying lighting on (a) Lena, (b) Cameraman, and (c) House images.

5.6 Overall Performance of All Matching Methods

The performance of the proposed matching method GOPM with the HD metric (HDGOPM) is measured and compared with other existing methods (traditional SSD, GOPM with the SSD metric). We have extended this experiment to more severe conditions in sections 6.1 and 6.2. Nine more standard test images are used in the simulation (total ten images). The occlusion patterns are also added with other patterns: Gaussian pattern and inverse of the Gaussian pattern. For the illumination, we have varied the global illumination of the second frame from -10% to -80%. For a matching method, the 5,393,132 motion vectors can be estimated. As a result, GOPM with HD can give the highest performance score for the motion estimation in artificial motion sequences, compared with the traditional SSD method. Notice that the threshold value of the method is 0.3. In addition, GOPM with SSD method has a slightly lower performance compared to the GOPM with HD.

Table 2 The average successful motion estimation rates of all matching methods and standard deviations are taken from 10 tested images, under 10 illumination conditions, 3 occlusion patterns, and 17 occlusion sizes.

Image SSD	*GOPM with*	*GOPM SSD*	*with the HD*
Mean (%)	4.462151	79.60105	80.56964
STD	12.50988	34.03117	36.37045

5.7 Test in Real Video Sequences

Finally, the proposed tracking method is tested on a real video sequence. In this experiment, a human face is the target for tracking. A facial region in the first frame of the video sequence is manually selected and used as a template for tracking. The video sequence includes four scenarios: normal situation, changing illuminations by casting a flashlight from various angles and distances, occlusion in which a foreign object covers part of the face, and the last scenario is a situation of both occlusion and varying illuminations (Fig. 11).

In order to measure the tracking efficiency of the GOPM with the HD metric (HDGOPM) qualitatively, we compare the tracking efficiency with other matching methods (the traditional SSD matching method and GOPM with the SSD metric). All matching techniques; traditional SSD, GOPM with the SSD metric, and GOPM with the HD metric ($\tau_1 = \tau_2 = 0.3$) perform the face tracking very well under a normal situation. However, in the case of changing illuminations, traditional SSD completely fails the tracking, while both GOPM with the SSD metric and GOPM with the HD metric work well. For the occlusion problem (and occlusion under varying illuminations), only GOPM with the HD metric (HDGOPM) can keep tracking the target, through almost the entire video sequence. Fig. 12 demonstrates the difference between the two tracking methods: GOPM with the SSD metric and GOPM with the HD metric. Fig. 12(a) shows that GOPM fails to detect the face under varying lighting conditions. Fig. 12(c) shows no prominent peak corresponding to the target object. On the other hand, the proposed method, GOPM with the HD metric (HDGOPM), successfully tracks the face as shown in Fig. 12(b). Fig. 12(d) clearly shows a prominent peak that indicates the position of the target with high confidence. It should be noted that the proposed method detects the target even under varying illuminations and with a significant amount of occlusion simultaneously.

Fig. 11. Four real image sequences; constant light without occlusion, irregular lighting without occlusion, constant lighting with occlusion, and irregular lighting with occlusion.

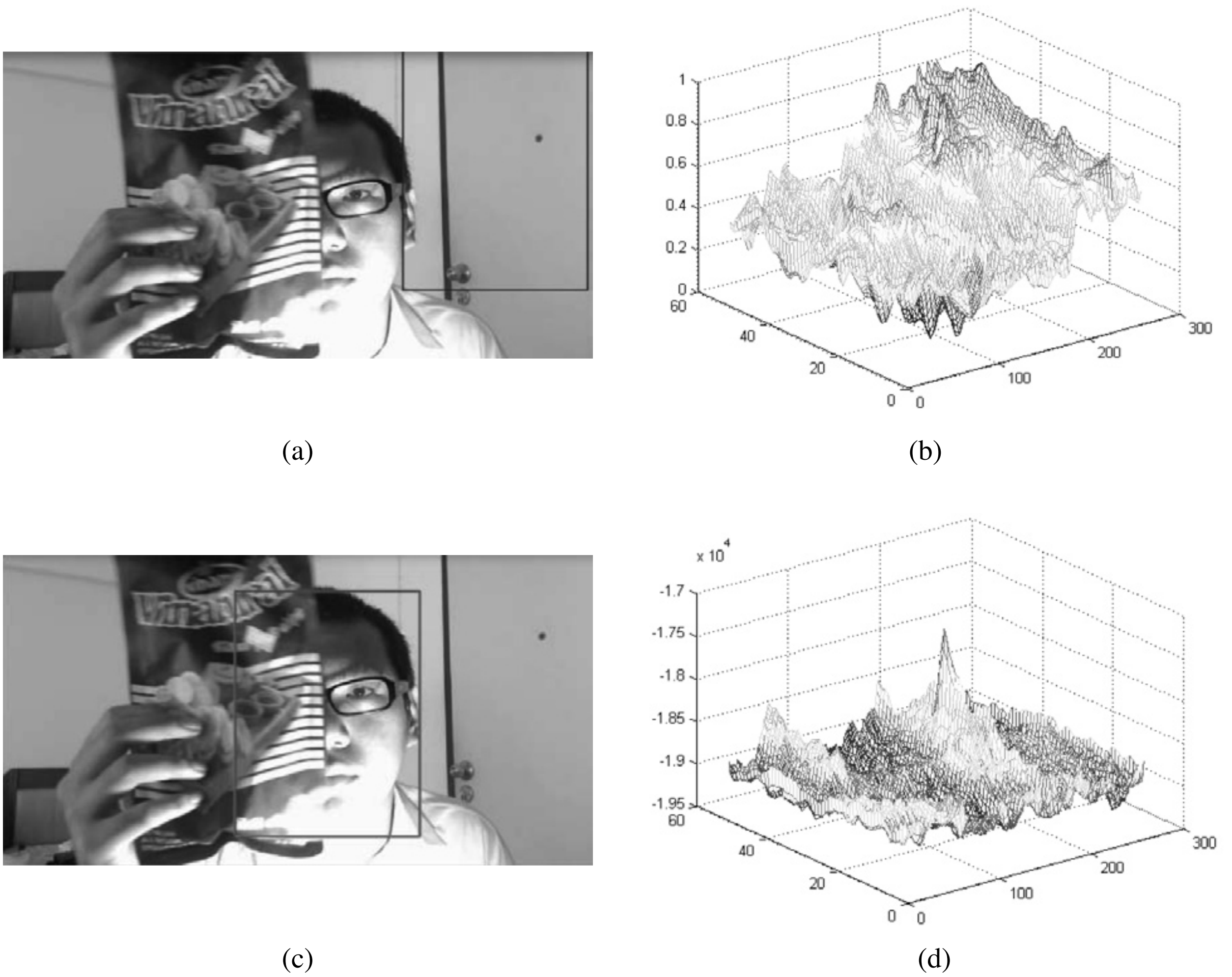


Fig. 12 Tracking results under irregular lighting with occlusion by (a) GOPM with SSD, (c) GOPM with HD, (b) and (d) the distributions of the corresponding similarity measurements

6. Conclusion

This paper presents a robust object tracking method GOPM with the HD metric (HDGOPM) that is based on unit gradient vectors (UGVs) and the Hamming distance (HD) metric. The technique performs well under varying illuminations and is also remarkably robust to the occlusion problem. Our simulation results show the advantages of the HD based gradient orientation pattern matching (GOPM) over existing matching techniques on both synthetic and real image sequences. Furthermore, we have compared three different strategies on how to convert the similarities between two corresponding UGVs to binary codes. Our study reveals that Strategy 3 (logical OR operation) is the best among the three approaches. We also investigate the effect of the threshold value to the performance of tracking. Experimental results show that the performance is not dependent on the threshold value, which means we may use a fixed threshold value rather than adjusting it according to input images. As a future work, we plan to work on the update of the template to handle varying appearances of a target.

7. Acknowledgements

This research is financially supported by Thailand Advanced Institute of Science and Technology (TAIST), National Science and Technology Development Agency (NSTDA), Tokyo Institute of Technology, and Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU) and the National Research University Project, Thailand Office of Higher Education Commission.

8. References

[1] Alper Yilmaz, Omar Javed, and Mubarak Shah, “Object Tracking: A Survey”, ACM Computing Surveys, Vol. 38, No. 4, Article 13, December 2006.

[2] Toshiaki Kondo and Waree Kongprawechnon, “A Block Matching Technique Using Unit Gradient Vectors”, MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN.

[3] Toshiaki Kondo and Waree Kongprawechnon, “A matching technique using gradient orientation patterns”, Thammasat Int. J. Sc. Tech., Vol. 14, No. 3, July-September 2009.

[4] Wattanit Hotrakool, Prarinya Siritanawan, and Toshiaki Kondo, “Real-time Gradient Orientation Pattern Matching”, International Conference on Embedded System and Information Technology, Chiang Mai, Thailand 2010.

[5] Wattanit Hotrakool, Prarinya Siritanawan, and Toshiaki Kondo, “A Real-time Eye-tracking Method using Time-varying Gradient Orientation Patterns”, In Proc. ECTI-CON, Thailand 2010.

[6] Prarinya Siritanawan and Toshiaki Kondo “Hamming Distance based Gradient Orientation Pattern Matching”, In Proc. International Symposium of Artificial life and Robotics 17th, January 2012 Ofir Pele and Michel Werman, “Robust Real-time Pattern Matching Using Bayesian Sequential Hypothesis Testing”, IEEE Transaction on Pattern Analysis and Machine Intelligence, August 2008.