HLAC Approach to Automatic Object Counting

Takumi Kobayashi† Tadaaki Hosaka† Shu Mimura‡ Takashi Hayashi§ Nobuyuki Otsu†

†National Institute of Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Japan

‡United Technologies Institute Co. Ltd.
1-1-1 Umezono, Tsukuba, Japan

§National Agricultural Research Center
3-1-1 Kannondai, Tsukuba, Japan

Abstract

Counting (identical) objects in images is a simple yet fundamental recognition task that requires exhaustive human effort. Automation of this task would reduce the human load significantly. In this paper, we propose a statistical method to automatically count objects in an image sequence by using Higher-order Local Auto-Correlation (HLAC) based image features and Multiple Regression Analysis (MRA). This method is based on a simple computation, which enables fast and automatic object counting in real time. We propose several methods that have different preprocessing and image features and conduct comparative experiments of counting objects (ducks in this paper) in images captured by outdoor monitoring cameras. The experimental results demonstrated the effectiveness of the proposed methods.

1. Introduction

Counting objects in images is necessary in various applications, such as counting people on roads or birds in the sky. It is a fundamental procedure for understanding a scene in which many objects exist. Although object counting is a simple task, exhaustive human efforts are required. In order to reduce the human loads associated with object counting, automation of the counting procedure using computer vision methods is desired.

The researches which are related to the task of counting objects have been based mainly on the detection of target objects [2, 11, 8]. In the object detection, detectors are constructed based on the model of predefined target patterns [8] or obtained by statistical learning [2, 11]. The detector generally consists of feature extraction from a small image region (called the detection window), e.g., HOG [2], and the classifier of the extracted features, e.g., SVM [10]. By sliding the detection window over the entire image, the detector responds at the positions where the target objects would be located, and the number of existing objects is then counted. Due to the exhaustive search over the entire image, the detection process is usually time-consuming. Moreover, for training the classifier in the detector, these methods require clipped images of the target objects including possible variabilities in terms of their appearances. Practically speaking, it is difficult to collect a large number of these images.

On the other hand, by using effective image features extracted from the entire image, even a simple statistical approach gives sufficiently favorable results. Otsu and Kurita [6] proposed Higher-order Local Auto-Correlation (HLAC) image features and applied them to the task of counting objects. Here, the number of objects in the image is estimated by means of multiple regression analysis (MRA), which is applied to HLAC feature vectors extracted from the given image. Unlike the object detection method described above, this method requires only pairs of an image and the correct number of objects in the image. Recently, Cubic Higher-order Auto-Correlation [3, 4] has been proposed for motion images and has been applied to the counting of moving objects in the motion images [3, 9]. These methods, however, have been applied only to controlled scenes, not to natural scenes.

In this paper, we propose a statistical method for counting identical objects in an image (sequence) by using an effective image feature extraction method based on HLAC and a statistical method by MRA, in a manner similar to [6]. Since the proposed method requires neither detection of object positions nor extraction of object figures, little computational time is required for automatic counting. We deal with natural color images captured by outdoor monitoring cameras, unlike in [6]. By taking into account the property

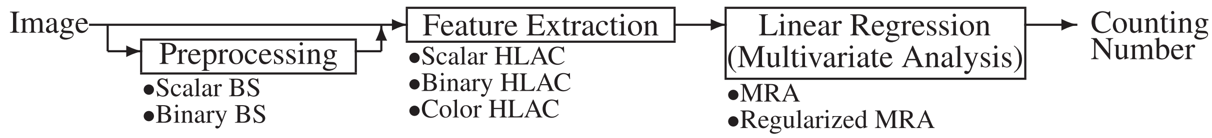


Figure 1. Overview of the proposed scheme. Each processing (denoted by boxes) has several choices (denoted by the items below the box).

of time sequencing in the image sequence, a regularization can be incorporated into the MRA. We propose several statistical methods that utilize different preprocessing and feature extraction in the configuration of the methods. In a comparative experiment involving the counting of ducks in natural color images, the proposed methods yield favorable results.

2. Proposed Method

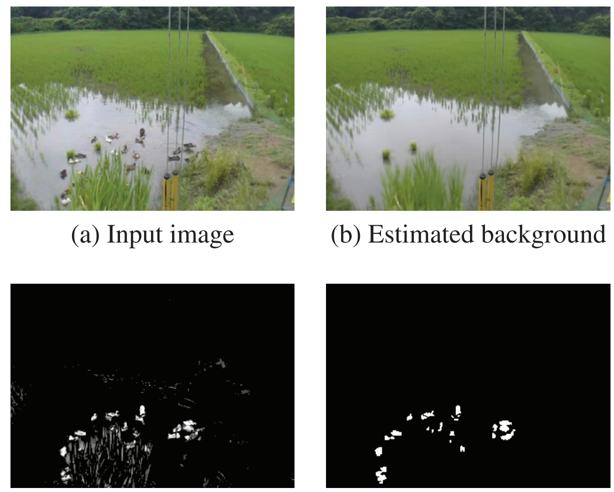
An overview of the proposed scheme is shown in Fig. 1. The proposed scheme consists primarily of three steps: preprocessing, feature extraction, and linear regression, the coefficients of which are obtained by multivariate analysis. Each step is described in detail in the following sections.

2.1. Preprocessing

Image sequences captured by monitoring cameras generally contain a static background. In this case, the method of background subtraction (BS) performs well and can eliminate the cluttered background. After such processing, the target objects are roughly cut out[1] and the image features of target objects can be effectively extracted, excluding the features of background. As preprocessing, we apply the following two types of the background subtraction method. One type is simple subtraction, in which the background image is estimated by calculating a median image from an image sequence at each pixel (Fig. 2(b)) and then is subtracted from each image. The pixels in the resulting image are the sum of the absolute difference values of RGB color channels, which form the gray-scale image, as shown in Fig. 2(c). We call this processing **Scalar Background Subtraction** (**Scalar BS**). The other type is a method similar to [7], which utilizes Graph Cuts [1] for definitely cutting out objects. The (foreground) object and backgrounds are represented by pixel values of 1 and 0, respectively, and the resulting image is binary (Fig. 2(d)). We call this **Binary Background Subtraction** (**Binary BS**).

Ordinary methods of object detection use these preprocessings. The proposed method based on HLAC-based image features and MRA, however, does not necessarily require these preprocessings, as described later, although, for the sake of comparison, they are applied in this paper.

[1] In the case that the preprocessing is applied, we assume that the target objects are moving, e.g., animals.



(a) Input image (b) Estimated background

(c) Scalar BS (d) Binary BS

Figure 2. Background subtraction

2.2. Feature Extraction

We employ the feature extraction methods based on Higher-order Local Auto-Correlation (HLAC) [6].

The N-th order HLAC is calculated by following auto-correlations:

$$R_N(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_N) = \sum_{\boldsymbol{r}} I(\boldsymbol{r}) I(\boldsymbol{r}+\boldsymbol{a}_1) \cdots I(\boldsymbol{r}+\boldsymbol{a}_N), \quad (1)$$

where I is the gray-scale image, $\boldsymbol{r} = (x, y)'$ (the dash denotes the transpose) is a position vector, and $\boldsymbol{a}_i$ are the displacement vectors. Since Eq.(1) takes so many forms by varying the parameter values $\boldsymbol{a}$ and N, these parameter values are restricted to vary as follows: $a_{ix}, a_{iy} \in \{\pm\Delta r, 0\}$ and $N \in \{0, 1, 2\}$. The displacement intervals (Δr) are the same in both the horizontal and vertical directions due to the isotropy of the image. The configuration $(\boldsymbol{r}, \boldsymbol{r}+\boldsymbol{a}_1, \cdots, \boldsymbol{r}+\boldsymbol{a}_N)$ is reduced to 35 (local mask) patterns, as shown in Fig. 3, by eliminating duplicates that arise from shifts. Thus, the HLAC feature for the gray-scale image is a 35-dimensional vector, called **Scalar HLAC**. In the case of binary images ($I(\boldsymbol{r}) \in \{0, 1\}$), the mask patterns are further reduced to 25 mask patterns (No.1$\sim$25 in Fig. 3). This is because, by considering $I(\boldsymbol{r}) = I(\boldsymbol{r})^2 = I(\boldsymbol{r})^3$, the mask patterns No.26$\sim$35 become equivalent to any of the other patterns. The HLAC feature for the binary image is a 25-dimensional vector, called **Binary HLAC**. Binary

HLAC extracts morphological characteristics in the image and has a linear relationship with the Euler number, as described in [6].

For color images, the Scalar HLAC extends to **Color HLAC** [5]. Let the RGB color image be $\boldsymbol{I} = (I_R, I_G, I_B)'$, in a manner similar to Eq.(1), the N-th order Color HLAC is calculated as follows:

$$\begin{aligned} & R_N(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_N, c_0, \cdots, c_N) \\ & \quad = \sum_{\boldsymbol{r}} I_{c_0}(\boldsymbol{r}) I_{c_1}(\boldsymbol{r}+\boldsymbol{a}_1) \cdots I_{c_N}(\boldsymbol{r}+\boldsymbol{a}_N), \end{aligned} \quad (2)$$

where c_i are color indices ($c_i \in \{R, G, B\}$) of $\boldsymbol{I}$. The parameter values are also restricted in the same manner as HLAC: the configurations of the displacement vectors are described by the 35 mask patterns shown in Fig. 3. The difference from Scalar HLAC is that color vectors $\boldsymbol{I}(\boldsymbol{r})$ are dealt with and all combinations of color indices (components) are taken in the auto-correlations of Eq.(2). The dimensionality of Color HLAC is 739, which is much higher than that of HLAC. Color HLAC has been successfully applied to an image retrieval task using human impression [5].

The HLAC-based features described above have the following desirable properties:

Shift-invariance to data: This is because the features are based on integral (summation). It is noteworthy that the shift-invariance renders the method *segmentation-free*.

Additivity for data: Suppose that regions A and B are disjoint ($A \cap B = \phi$) in an image. Then, the feature vector from the entire image is given as

$$R_{\text{whole}} = \sum_{\boldsymbol{r} \in (A \cup B)} g(\boldsymbol{r}) \approx \sum_{\boldsymbol{r} \in A} g(\boldsymbol{r}) + \sum_{\boldsymbol{r} \in B} g(\boldsymbol{r}) = R_A + R_B,$$

where $g(\boldsymbol{r}) = I(\boldsymbol{r}) I(\boldsymbol{r}+\boldsymbol{a}_1) \cdots I(\boldsymbol{r}+\boldsymbol{a}_N)$. This holds because auto-correlations are almost limited to each region (A or B) due to their locality. This property also makes it possible to simultaneously identify multiple objects [3].

These properties are well-suited to subsequent linear methods, and, in particular, are desirable for the task of counting object by eliminating the need for detecting and extracting objects from cluttered backgrounds.

2.3. Multivariate Analysis

In the training phase, the pairs of the feature vectors $\boldsymbol{x}_i$ and the correct numbers of target objects (teacher signals) c_i for the i-th image are given. We apply **Multiple Regression Analysis (MRA)**, which determines the optimal linear coefficients $\boldsymbol{a}$, to estimate c from $\boldsymbol{x}$: $c \approx \boldsymbol{a}'\boldsymbol{x} + b = \hat{\boldsymbol{a}}'\hat{\boldsymbol{x}}$, where b is constant, $\hat{\boldsymbol{a}} = [\boldsymbol{a}', b]'$, $\hat{\boldsymbol{x}} = [\boldsymbol{x}', 1]'$. $\hat{\boldsymbol{a}}$ is obtained

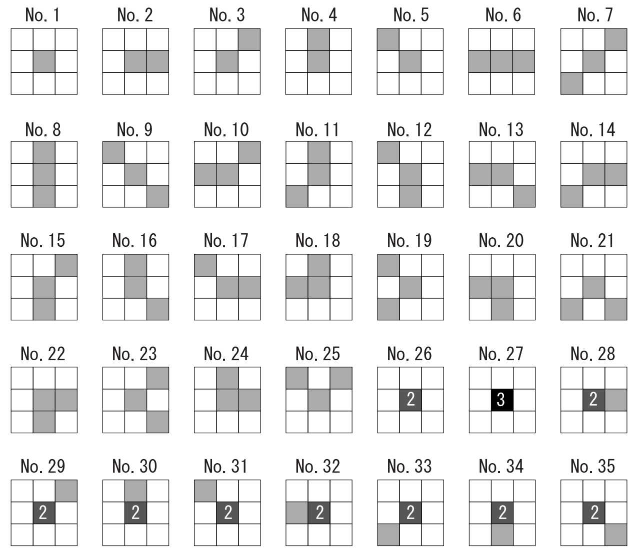


Figure 3. Mask patterns. The numbers in the cells indicate duplicated positions for auto-correlations.

by minimizing the following regression error:

$$\begin{aligned} \min_{\boldsymbol{a}, b} \sum_i ||c_i - \boldsymbol{a}'\boldsymbol{x}_i - b||^2 &= \min_{\hat{\boldsymbol{a}}} \sum_i ||c_i - \hat{\boldsymbol{a}}'\hat{\boldsymbol{x}}_i||^2 \\ &\Rightarrow \hat{\boldsymbol{a}} = (\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}')^{-1}\hat{\boldsymbol{X}}\boldsymbol{c} \end{aligned} \quad (3)$$

where $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_N], \boldsymbol{c} = [c_1, \cdots, c_N]'$. Given the image feature $\boldsymbol{x}$ extracted from the test image, the number of objects in the image can be estimated by $\tilde{c} = \boldsymbol{a}'\boldsymbol{x} + b$.

In this study, we deal with successive images captured by monitoring cameras. For these images, which have the property of time sequencing, the estimation results for successive frames should be smooth along the time axis. The MRA incorporating this constraint can be defined as follows:

$$\min_{\boldsymbol{a}, b} \sum_i ||c_i - \boldsymbol{a}'\boldsymbol{x}_i - b||^2 + \lambda ||\boldsymbol{a}'(\boldsymbol{x}_i - \boldsymbol{x}_{i+1})||^2 \quad (4)$$

$$\Rightarrow \hat{\boldsymbol{a}} = \left\{ (1+2\lambda)\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}' - \lambda \sum_i (\hat{\boldsymbol{x}}_i\hat{\boldsymbol{x}}'_{i+1} + \hat{\boldsymbol{x}}_{i+1}\hat{\boldsymbol{x}}'_i) \right\}^{-1} \hat{\boldsymbol{X}}\boldsymbol{c}$$

where λ is the balancing parameter. We call this the **Regularized MRA (R-MRA)**. The second term in Eq.(4) plays the role of a penalty for drastic changes of the estimated numbers and possibly smoothes (or stabilizes) the results. It is also expected that regularization enhances the generalization performance by avoiding overfitting of the model.

Due to the property of additivity in HLAC, the MRA can be applied to the estimation of the number of objects without even the preprocessing of the background subtraction.

Table 1. Configurations of the proposed methods

	Preprocessing	Feature	Analysis
Col MRA	None	Color HLAC	MRA
Col R-MRA	None	Color HLAC	R-MRA
Sca MRA	Scalar BS	Scalar HLAC	MRA
Sca R-MRA	Scalar BS	Scalar HLAC	R-MRA
Bin MRA	Binary BS	Binary HLAC	MRA
Bin R-MRA	Binary BS	Binary HLAC	R-MRA

Table 2. Mean Error in each sequence

Method	Seq1	Seq2	Seq3	Seq4	Seq5	Total
Col MRA	3.49	4.53	**4.69**	2.73	4.01	3.89
Col R-MRA	3.46	4.53	4.79	**2.65**	**3.97**	**3.88**
Sca MRA	5.54	5.25	5.75	6.48	9.36	6.48
Sca R-MRA	5.26	5.48	5.70	6.59	9.16	6.44
Bin MRA	2.86	5.47	6.36	7.85	5.99	5.70
Bin R-MRA	**2.17**	**3.59**	6.71	8.43	5.94	5.37

In the case that the objects are cut out by the preprocessing, the constant term b derived from the constant background is not required in MRA, i.e., $b = 0$.

2.4. Configuration of the Proposed Methods

In each step described in Fig. 1, we have several choices, as described above. The configurations of the six proposed methods are shown in Table 1. These methods have different preprocessing and feature extraction methods for the two types of MRA. **Col MRA** is similar to [6], which extends from HLAC to Color HLAC. **Bin MRA** also utilizes the method of [6] by binarizing the input color images. While **Sca MRA** and **Bin MRA** utilize BS methods, **Col MRA** does not include BS and is purely segmentation-free.

3. Experimental Result

We apply the proposed methods to the task of counting ducks (identical objects) outdoors (Fig. 4). The images were successively captured (at approximately 0.5 fps) by cameras monitoring rice paddies, and five image sequences were obtained. The number of ducks is counted (estimated) at each frame. For evaluating the performances of the proposed methods, a leave-one-out scheme is applied to these sequences. The linear coefficients are obtained by MRA applied to four sequences and the methods are tested on the remaining sequence, and the mean error (of the number of ducks) is then calculated.

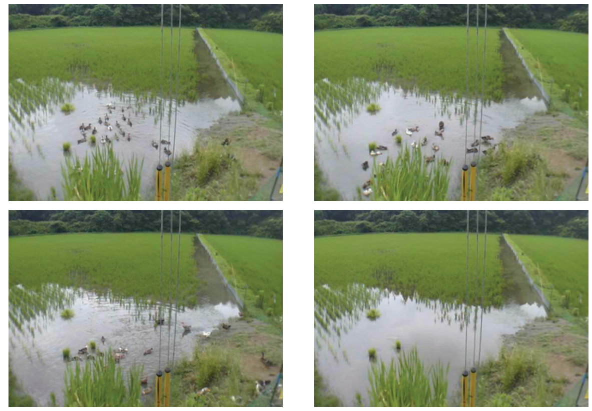


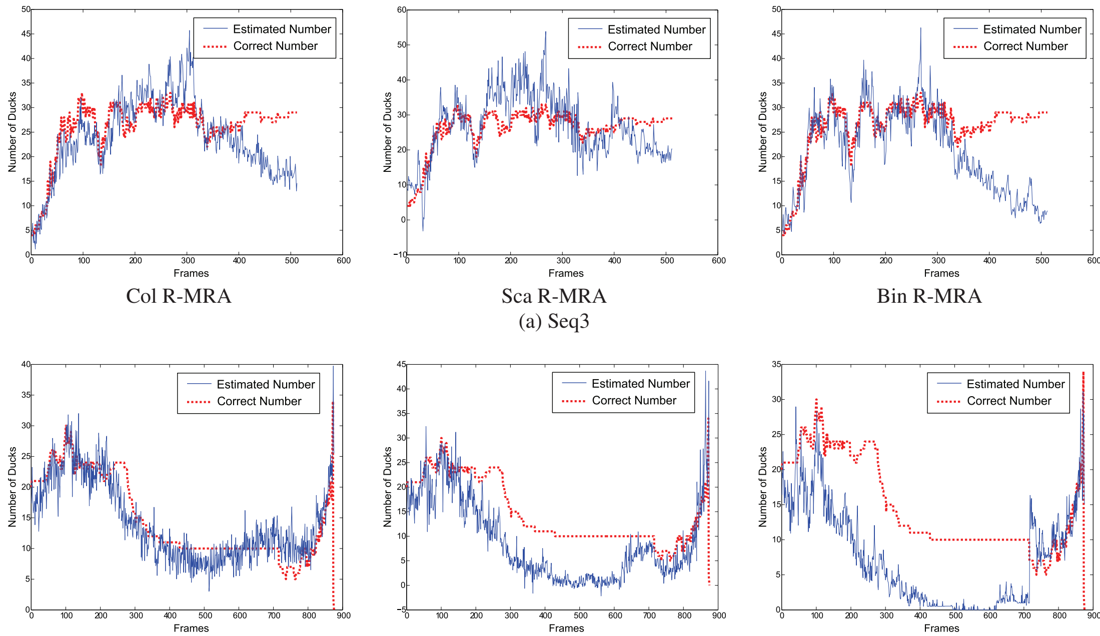
Figure 4. Example images

The results are shown in Table 2. The average number of ducks is 17.5. In all of the methods, the regularization in R-MRA slightly improves the performance of MRA by stabilizing the results of the estimation.

The method combining Scalar Background Subtraction and Scalar HLAC was the least effective. In this method, the objects are roughly cut out, but the resulting pixel values within object regions in the processed images are affected by background pixel values due to simple subtraction. This results in large variabilities of features of objects (ducks), and the linear regression cannot capture these large variations.

The method of Binary Background Subtraction and Binary HLAC avoid this problem by definitely extracting objects and constructing binary images (Fig. 2(d)). This yields favorable results on several sequences, such as Seq 1 and Seq 2, as shown in Table 2. However, this sophisticated preprocessing has significant effects on the results. If the objects were missed by this preprocessing, they could not be recovered in subsequent processes. Thus, this method does not work well for images in which background subtraction fails to cut out objects (e.g., Seq 4) and the Binary HLAC can no longer extract efficient features of objects. Moreover, BS methods extract noises derived from, e.g., swinging grasses, etc.

The method of Color HLAC yielded favorable results for all sequences and appears to be the most effective method, despite being quite simple in that it does not require preprocessing. The properties of additivity and shift-invariance in (Color) HLAC features are suited to linear regression of MRA such that there is no need for preprocessing such as object detection or even background subtraction, which would affect the results. The reason for this is as follows. Let the feature vectors of the object, constant background, and the other variations be $\boldsymbol{f}, \boldsymbol{b}, \boldsymbol{n}_i$, and let the number of the objects be c_i for the i-th image. Due to the additivity of HLAC features, the feature vector $\boldsymbol{x}_i$ extracted from the (i-th) whole image can be decomposed as $\boldsymbol{x}_i = c_i\boldsymbol{f} + \boldsymbol{n}_i + \boldsymbol{b}$.



Col R-MRA Sca R-MRA Bin R-MRA

(a) Seq3

Col R-MRA Sca R-MRA Bin R-MRA

(b) Seq4

Figure 5. Estimated number of objects on the image sequences of Seq3 (a) and Seq4 (b).

Suppose that the vector $\tilde{\boldsymbol{a}}$ is orthogonal to all the variations $\boldsymbol{n}_i$, i.e., $\tilde{\boldsymbol{a}} \perp \boldsymbol{n}_i \ \forall i$. The counting number c_i can be retrieved by $c_i = \frac{\tilde{\boldsymbol{a}}'(\boldsymbol{x}_i - \boldsymbol{b})}{\tilde{\boldsymbol{a}}'\boldsymbol{f}} = \boldsymbol{a}'(\boldsymbol{x}_i - \boldsymbol{b})$, which is the same formulation as the linear regression model in (3). Even if we do not know the actual object features $\boldsymbol{f}$, the vectors $\boldsymbol{a}, \boldsymbol{b}$ can be statistically obtained by MRA from many sample pairs of the features $\boldsymbol{x}_i$ and the correct numbers c_i. Thus, the linear models of MRA together with the additivity property of Color HLAC features is quite effective in terms of not only the accuracy of the estimation, but also computational cost.

Finally, we present examples of the estimated numbers on several image sequences in Fig. 5. As shown in Fig. 5(b), the methods of **Sca R-MRA** and **Bin R-MRA** utilizing BS preprocessing failed, and the estimation results were far from the ground truth.

4. Conclusion

We have proposed statistical methods for counting objects in image sequences. The methods comprise HLAC-based image features and Multiple Regression Analysis (MRA). The properties of *additivity* and *shift-invariance* in HLAC are well suited to the linear model of MRA, and thus do not require complicated processings, such as the detection of objects and their locations or the extraction of object figures. By taking into account the continuity of time sequencing, we also propose Regularized MRA. In the comparative experiments of counting ducks in the images captured by outdoor monitoring cameras, the results demonstrated the effectiveness of the proposed methods. In particular, the method combining Color HLAC and Regularized MRA yielded the most favorable results. Unlike the conventional methods, the method are completely segmentation-free.

In this paper, we assume that the target objects are identical throughout image sequences. The proposed methods, however, can naturally deal with several kinds of objects to be counted by converting the (scalar) number c to the vector $\boldsymbol{c}$ consisting of the numbers of each kind of object.

References

[1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 20–25, 2005.

[3] T. Kobayashi and N. Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In *International Conference on Pattern Recognition*, pages 741–744, 2004.

[4] T. Kobayashi and N. Otsu. A three-way auto-correlation based approach to human identification by gait. In *IEEE Workshop on Visual Surveillance*, pages 185–192, 2006.

[5] T. Kurita, S. Kato, I. Fukuda, and A. Itakura. Sense retrieval on a image database of full color paintings (in japanese). *Information Processing Society of Japan*, 33(11):1373–1383, 1992.

[6] N. Otsu and T. Kurita. A new scheme for practical flexible and intelligent vision systems. In *IAPR Workshop on Computer Vision*, 1988.

[7] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *Pattern Analysis and Machine Intelligence*, 27(11):1778–1792, 2005.

[8] T. Shimada, K. Kato, A. Kamikouchi, and K. Ito. Analysis of the distribution of the brain cells of the fruit fly by an automatic cell counting algorithm. *Physica A*, 350:144–149, 2004.

[9] Y. Shimohata and N. Otsu. Real-time and simultaneous recognition of multiple moving objects using cubic higher-order local auto-correlation. In *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 49–52, 2008.

[10] V. Vapnik, editor. *Statistical Learning Theory*. Wiley, 1998.

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 511–518, 2001.