Higher-order Co-occurrence Features based on Discriminative Co-clusters for Image Classification

Takumi Kobayashi
takumi.kobayashi@aist.go.jp

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

Abstract

We propose a method to extract image features based on effective higher-order co-occurrences. The proposed method constructs the *co-clusters* to discriminatively quantize joint primitive quantitative data, such as pair-wise pixel intensities, unlike the standard co-occurrence methods that utilize simple clusters trained in an unsupervised manner for quantizing point-wise data. The discriminative co-clusters effectively exploit the co-occurrence characteristics even by a fewer number of cluster components, resulting in low-dimensional co-occurrence features. By taking advantage of those discriminative co-clusters, the co-occurrence features can be extended to the higher-order co-occurrence features of feasible dimensionality. The higher-order co-occurrence captures richer information in image textures by extracting relationships in multiplets more than only doublets (pairs). In the experiments on image classifications for cancer cells and pedestrians, the proposed method exhibits favorable performances compared to the other methods, even to the standard co-occurrence based methods.

1 Introduction

Image classification is important to tackle a variety of real-world problems such as computer aided diagnosis and image surveillance, and much research effort has been made in the computer vision community. A feature extraction is especially a fundamental procedure to improve the performances of the image classification. It is necessary to extract characteristics of target objects and textures with retaining robustness to irrelevant variations derived from environmental changes, such as changes in illumination or target position.

Histogram-based feature extraction methods have exhibited promising performances, *e.g.*, SIFT [11], HOG [5], color histogram [15, 18] and bag-of-feature (BoF) [4]. Those methods statistically extract image features by measuring occurrences of the *qualitative* data (referred to as *symbols* in this paper) in the form of histograms; examples of the symbols are gradient orientation bins in SIFT and HOG, indexed colors in the color histogram and visual words in BoF. The statistical features are robust to noises and they are fed into subsequent classification methods, such as SVM [19], for accomplishing the image classification. Beyond the histogram-based methods considering occurrences, co-occurrence feature extraction methods have also attracted keen attentions thanks to the superior performances [6, 7, 8, 9, 10, 17, 23, 24]. The methods statistically describe the image by using

co-occurrences of symbols on *pair*-wise pixels, while the histogram-based methods mentioned above are based only on the *point*-wise occurrence. The co-occurrence can capture the richer information, *i.e.*, joint information of the symbols, and the occurrence features are regarded as the marginal ones of the co-occurrence features from the probabilistic viewpoint. The co-occurrence features that exploit the local pixel pairs also retain the shift invariance as in the histogram-based features, which is favorable to image classifications.

For extracting the co-occurrences, it is common to transform the *quantitative* data into *qualitative* data (symbols) by means of quantization (clustering) at first, as is the case with the histogram-based methods; for example, continuous gradient orientation is coded into orientation bins [8, 17], RGB colors are indexed [7] and local features are categorized into visual words [10, 24], although a few works [1, 13] attempt to directly code joint pairs of local features for BoF. The quantization process taken over from the histogram-based methods is given a priori in an ad-hoc manner based on prior knowledge regarding the point-wise statistics of the quantitative data, not pair-wise data. Therefore, the obtained qualitative symbols are not necessarily suitable to characterize the co-occurrence. And, the co-occurrence features have been computed based only on pair-wise symbols and higher-order co-occurrences beyond pair-wise ones have been rarely considered so far due to exponential increase of the dimensionality; even 10 types of symbol produce 10^4 dimensionality for the quadruplet co-occurrence.

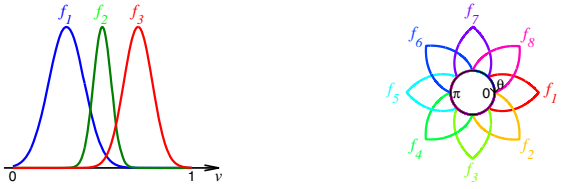
In this paper, we propose a method to extract higher-order co-occurrence image features. The proposed method is built upon the *co-clusters* discriminatively quantizing *pair-wise* quantitative data, in contrast to the standard methods that utilize simple clusters of point-wise data trained in an unsupervised manner. The discriminative co-clusters directly capture the statistical characteristics, *i.e.*, co-occurrence, of pair-wise data, and effective co-occurrence features are extracted by using even a small number of the co-clusters, which results in low dimensionality. Thus, we can develop the higher-order co-occurrence feature of feasible dimensionality based on co-occurrences of quadruplets which are pairs of pair-wise data represented by the discriminative co-clusters. The higher-order co-occurrences exploit richer information in image textures by taking into account of higher-order relationships in multiplets more than doublets (pairs) and contribute to improve the performance of image classifications.

2 Discriminative co-clusters for co-occurrence features

2.1 Standard co-occurrence features

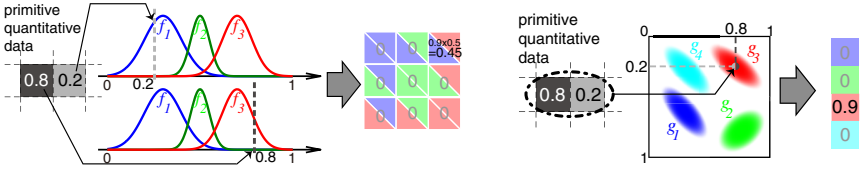
Co-occurrence features first appeared in the work by Haralick *et al.* [6] which characterizes image textures by gray-level co-occurrence matrix (GLCM). Recently, the standard co-occurrence features are founded on the qualitative data (symbols); gradient orientation bins [8, 17], indexed colors [7] and visual words [10, 24]. These methods first cluster (quantize) primitive *quantitative* data, *e.g.*, gradient orientations, RGB colors and local features [11], into those symbols and then measure the co-occurrences among them. The co-occurrences of the primitive quantitative data as in GLCM [6] are considered to be sensitive to the irrelevant variations derived from noise. Through the process of information reduction by quantizing the quantitative data, the resultant features are robust against such variations, improving the classification performances.

In summary, we mathematically describe the process to construct the standard co-occurrence



(a) on pixel intensities v (b) on gradient orientations θ (cyclic)

Figure 1: Examples of cluster assignment functions f_i.



(a) Standard co-occurrence using f_i (b) Proposed co-occurrence using g_k

Figure 2: Procedures to extract co-occurrence features.

features as follows. Let $\mathcal{R}$ be the quantitative data space[1], $x_{\boldsymbol{p}} \in \mathcal{R}$ be the quantitative data at pixel position $\boldsymbol{p}$ in an image plane. We introduce the function $f_i(x) : \mathcal{R} \to \Re_+$ to assign $x \in \mathcal{R}$ with the membership to the i-th cluster ($i = 1, \cdots, C$), as shown in Fig. 1. Note that we generally consider the soft assignment by f_i, not only hard assignment $f_i : \mathcal{R} \to \{0,1\}$. The cluster assignment functions f_i are usually determined a priori or in an unsupervised manner such as by applying k-means or EM method. The co-occurrence features are defined by

$$\boldsymbol{M} = \left\{ \sum_{\{\boldsymbol{p},\boldsymbol{q}\} \in \mathbb{N}} \omega(\boldsymbol{p},\boldsymbol{q}) f_i(x_{\boldsymbol{p}}) f_j(x_{\boldsymbol{q}}) \right\}_{i,j=1,\cdots,C} \in \Re^{C \times C}, \tag{1}$$

where $\mathbb{N}$ indicates the set of local neighbor pairs; *e.g.*, $\mathbb{N} = \{(\boldsymbol{p},\boldsymbol{q}) \big| \|\boldsymbol{p} - \boldsymbol{q}\| = \Delta\}$ where Δ is the displacement interval for local neighborhoods. In this paper, unlike the legacy co-occurrences [6, 7], we introduce the weighting function ω on pixel pairs, the actual form of which is practically defined depending on the task (see Sec.4) as in [8].

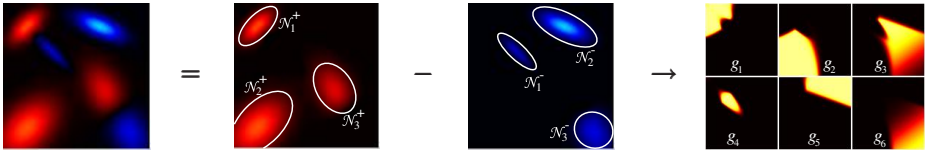
2.2 Discriminative co-clustering

We consider the following general form for extracting the co-occurrence features:

$$\boldsymbol{M} = \left\{ \sum_{\{\boldsymbol{p},\boldsymbol{q}\} \in \mathbb{N}} \omega(\boldsymbol{p},\boldsymbol{q}) g_k(x_{\boldsymbol{p}}, x_{\boldsymbol{q}}) \right\}_{k=1,\cdots,D} \in \Re^D, \tag{2}$$

where we introduce the function $g_k(x_{\boldsymbol{p}}, x_{\boldsymbol{q}}) : \mathcal{R} \times \mathcal{R} \to \Re_+$ to assign the pair $(x_{\boldsymbol{p}}, x_{\boldsymbol{q}})$ with the k-th cluster ($k = 1, \cdots, D$) in the joint space $\mathcal{R} \times \mathcal{R}$, called *co-cluster*. The formulation (1) is the special case of (2) since (2) corresponds to (1) in the case that the co-clusters can be factorized by $g_k(x_{\boldsymbol{p}}, x_{\boldsymbol{q}}) = f_i(x_{\boldsymbol{p}}) f_j(x_{\boldsymbol{q}})$ with $D = C^2$. The formulation (1) using the factorized cluster assignment functions assumes that the primitives $x_{\boldsymbol{p}}, x_{\boldsymbol{q}}$ act *independently*, although those local neighborhoods actually work on each other. The general form (2) naturally copes

[1] For example, $\mathcal{R} = [0,1]$ for pixel intensities, and $\mathcal{R} = [0, 2\pi)$ for gradient orientations.



classifier weight $\boldsymbol{W}$ positive weight $\boldsymbol{W}^+$ negative weight $\boldsymbol{W}^-$ co-clusters (posteriors) g_k

Figure 3: Construction of discriminative co-clusters g_k.

with such mutual interferences by using the co-cluster assignment function g_k in the joint space $\mathcal{R}\times\mathcal{R}$. The primitive co-occurrence on $\mathcal{R}\times\mathcal{R}$ of highly correlated pairs can not be efficiently characterized by using the factorized functions f_i, while even small number of the joint functions g_k can exploit it effectively. The procedure to extract the proposed co-occurrence features by using g_k is illustrated in Fig. 2 with comparison to that of the standard co-occurrences (Sec.2.1).

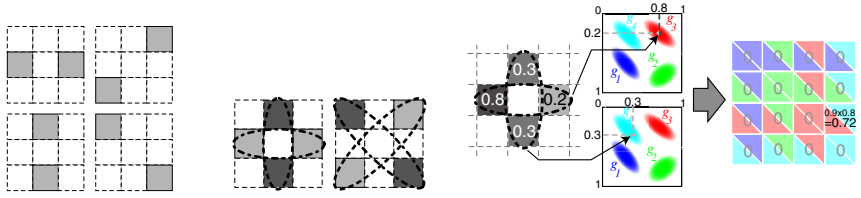
We determine the function g_k in a discriminative manner. Suppose a two-class problem of images I_n with the class label $y_n \in \{+1,-1\}$. From the image I_n, we first extract primitive co-occurrence features $\tilde{\boldsymbol{M}}_n$ on $\mathcal{R}\times\mathcal{R}$ as in GLCM [6]; in practice, the space $\mathcal{R}$ which is usually continuous is finely partitioned into (large number of) L bins, resulting in $\tilde{\boldsymbol{M}}_n \in \Re^{L\times L}$. Then, the *linear* SVM [19] is applied to those $(\tilde{\boldsymbol{M}}_n, y_n)$ in order to produce the classifier weight $\boldsymbol{W} = \sum_n \beta_n y_n \tilde{\boldsymbol{M}}_n$ on $\mathcal{R}\times\mathcal{R}$, actually $\boldsymbol{W} \in \Re^{L\times L}$, where β_n are the Lagrange multipliers in SVM. The classifier weight exploits the discriminative information: the positive weights in $\boldsymbol{W}$ contribute to '+1' class, while the negative ones to '-1' class. Finally, we perform clustering on the weight matrix $\boldsymbol{W}$ to produce the co-cluster assignment function g_k which is determined as the membership function to the k-th (co-)cluster on $\mathcal{R}\times\mathcal{R}$. We separately treat the weight $\boldsymbol{W}$ in terms of its sign (positive/negative) as the positive weight $\boldsymbol{W}^+ = \max(\boldsymbol{W}, 0)$ and the negative $\boldsymbol{W}^- = \max(-\boldsymbol{W}, 0)$, $\boldsymbol{W} = \boldsymbol{W}^+ - \boldsymbol{W}^-$, and apply the clustering method to those respective weights as follows.

Though any kinds of clustering methods like mean shift [3] and quick shift [20] are applicable, in this study we adopt the EM method based on the mixture of Gaussian distribution because the method provides soft membership to the clusters with a free parameter of the number of clusters. Soft assignment would reduce the quantization errors, improving the classification performances [16]. The parameter for the number of clusters easily controls the dimensionality of the resultant co-occurrence features. Whilst, the other methods, mean shift [3] and quick shift [20], work for hard segmentation with the bandwidth parameter instead of the number of clusters. By applying the EM method [22] to the respective weights $\boldsymbol{W}^{+/-}$ regarded as probability densities, we obtain the cluster component (Gaussian) functions $\mathcal{N}_k^{+/-}$ with the prior weights $\alpha_k^{+/-}$, which are gathered into the set of $\{\mathcal{N}_k, \alpha_k, \rho_k\}_{k=1,\cdots,D}$ where $\rho_k \in \{\rho^+, \rho^-\}$ indicates the prior either of positive ρ^+ or negative ρ^-. The function g_k is finally determined by

$$g_k(x_1,x_2) = \frac{\rho_k \alpha_k \mathcal{N}_k(x_1,x_2)}{\sum_k \rho_k \alpha_k \mathcal{N}_k(x_1,x_2)}, \quad \forall x_1, x_2 \in \mathcal{R}, \tag{3}$$

which is the posterior probability at (x_1,x_2), resulting in the normalized g_k: $\sum_k g_k(x_1,x_2) = 1$. Clusters in the positive and negative weights are integrated via the priors $\rho^{+/-}$ and we set them to $\rho^+ = \rho^- = 0.5$ in this study. The above-mentioned procedure to construct the discriminative co-clusters g_k is illustrated in Fig. 3.

In the standard method (Sec.2.1), the cluster assignment functions f_i are given simply by clustering point-wise quantitative data samples, *i.e.*, (point-wise) primitive occurrences, in an



(a) doublets (pairs) (b) *cross* quadruplet (c) higher-order co-occurrence

Figure 4: Higher-order co-occurrence.

unsupervised manner. The primitive occurrences, however, do not have enough discriminative power. On the other hand, the primitive co-occurrences are more discriminative than the occurrences, and thus the well-localized classifier weights are obtained. Our contributions are 1) to consider co-occurrences from the top for quantization and 2) to discriminatively construct co-clusters based on the (initial) classification weight by the *linear* SVM. A few BoF methods [1, 13] consider the clustering of joint pairs of local features, but the clusters are constructed in an unsupervised manner.

3 Higher-order co-occurrence features

The co-cluster assignment functions g_k discriminatively characterize joint (pair-wise) quantitative data, and then we obtain the co-occurrence features (2) of usually lower dimensionality D than C^2 of the standard features (1) using the factorized functions $f_i f_j$. This is because the function g_k can naturally cope with the joint relationship of correlated data at once without assuming factorization $f_i f_j$. Based on such fact, we further develop the higher-order co-occurrence features on the multiplets more than doublets (pairs). In this paper, we consider the co-occurrence of *quadruplets* which are pairs of pair-wise data.

The proposed higher-order co-occurrence features are defined by

$$\boldsymbol{H} = \left\{ \sum_{\{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r},\boldsymbol{s}\}\in\mathbb{Q}} \omega(\boldsymbol{p},\boldsymbol{q},\boldsymbol{r},\boldsymbol{s}) g_k(x_{\boldsymbol{p}},x_{\boldsymbol{q}}) g_l(x_{\boldsymbol{r}},x_{\boldsymbol{s}}) \right\}_{k,l=1,\cdots,D} \in \Re^{D\times D}, \tag{4}$$

where $\mathbb{Q}$ indicates the quadruplets. In this higher-order co-occurrence, it is important how to determine the quadruplets $\mathbb{Q}$, forms of which could be combinatorially increased. Co-occurrences are based on pairs which are oriented in various directions as shown in Fig. 4a, and we configure the quadruplets, the pairs of pairs, in the form of *cross* as shown in Fig. 4b in order to extract diverse characteristics in image textures; the pairs in the cross are maximally (orthogonally) separated.

The procedure to extract the higher-order co-occurrence features is illustrated in Fig. 4c. The formulation (4) is similar but one-order higher than that of the standard co-occurrence (1) (Fig. 2a) by replacing the functions f_i, f_j with g_k, g_l, point-wise data $x_{\boldsymbol{p}}, x_{\boldsymbol{q}}$ with pair-wise data $(x_{\boldsymbol{p}}, x_{\boldsymbol{q}}), (x_{\boldsymbol{r}}, x_{\boldsymbol{s}})$, and thus doublets (pairs) of elements $(x_{\boldsymbol{p}}, x_{\boldsymbol{q}})$ with the quadruplets $(x_{\boldsymbol{p}}, x_{\boldsymbol{q}}, x_{\boldsymbol{r}}, x_{\boldsymbol{s}})$. In the proposed features, the co-occurrences are computed by the histogram on g_k in (2) and the higher-order co-occurrences are given by the co-occurrence on $g_k g_l$ in (4). In case that we use the standard co-occurrences $f_i f_j$ for this higher-order co-occurrence, the dimensionality of the resultant feature is significantly large, C^4, which is infeasible from the practical viewpoint. In addition, the higher dimensionality would cause over-fitting to the training samples, degrading the performance. In the proposed method, by

utilizing smaller number of the discriminative co-clusters, the dimensionality of the higher-order co-occurrence feature is feasibly low with capturing the relationships in the quadruplets to extract the discriminative higher-order information in the image textures. Note that the features in (4) are finally fed into the classification method, apart from the initial classification by linear SVM in constructing co-clusters g_k (Sec.2.2).

4 Experimental results

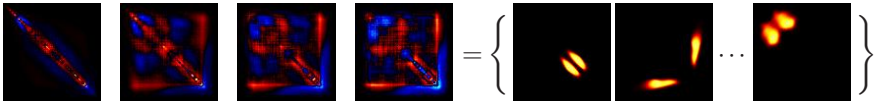
We apply the proposed method to two image classification tasks: cancer detection and pedestrian detection which result in binary (two class) classifications as cancer vs. non-cancer and pedestrian vs. non-pedestrian. For discriminative co-clusters (Sec.2.2), we apply the EM method to the positive weight $\boldsymbol{W}^+$ and negative $\boldsymbol{W}^-$ with the half number of D, respectively; *e.g.*, in the case of $D = 20$, we obtain respective 10 co-clusters from positive and negative weights. In these experiments, the (higher-order) co-occurrence feature matrices $\boldsymbol{M}$ ($\boldsymbol{H}$) are unfolded into vector forms normalized in unit L_2 norm and classification is finally performed by applying linear SVM [19] to the feature vectors.

4.1 Cancer detection

First, for computer-aided diagnosis (CAD), we conducted the experiment on cancer detection by using a high-resolution biopsy image including stomach cancer cells which metastasized to lymph nodes. The biopsy image is obtained by H&E staining and tissues are stained by violet blue or light red as shown in Fig. 7. The original huge image of 21168×14992 pixels is split up into six sub-images of the identical size (7056×7496) and the pixel-wise labels indicating cancerated regions are correspondingly given by the expert pathologists. The features are extracted on the running windows of 200×200 pixels shifted by 100 pixels; namely, such patches of 200×200 pixels are treated as the input images I_n with the label $y_n = \mathrm{sign}(E_{\boldsymbol{p} \in patch} y_{\boldsymbol{p}})$, the sign of the mean label within the patch. We conducted six-fold cross validations by using those six sub-images and measured the performances of equal error rate (EER).

A preprocessing is applied to extract the foreground (lymph nodes) region by eliminating background pixels based on the prior knowledge as in [23]. The background where tissues are absent are not stained, resulting in white colors, and such regions are irrelevant to cancer detection. Since the biopsy images are stained mainly by red and blue colors via H&E, the remaining green channel is useful only to distinguish the background regions by simple thresholding: the threshold value is 200 in this study. The foreground/background at the pixel $\boldsymbol{p}$ is indicated by $l_f(\boldsymbol{p}) = 1/0$, and for feature extraction, the weighting functions are defined as $\omega(\boldsymbol{p}, \boldsymbol{q}) = l_f(\boldsymbol{p}) l_f(\boldsymbol{q})$ in (2) and $\omega(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}, \boldsymbol{s}) = l_f(\boldsymbol{p}) l_f(\boldsymbol{q}) l_f(\boldsymbol{r}) l_f(\boldsymbol{s})$ in (4), which makes us compute the features only on the foreground regions. Then, the R-B channels are reduced into gray-scale by applying PCA to the foreground pixels, and the gray-scale values are finely partitioned into 100 levels ($L = 100$ in Sec.2.2).

The neighborhoods $\mathbb{N}$ for computing the co-occurrence are defined by $\mathbb{N} = \{(\boldsymbol{p}, \boldsymbol{q}) \big| \|\boldsymbol{p} - \boldsymbol{q}\| = \Delta\}$ where Δ indicates the displacement interval. Since the cells are arbitrarily oriented in the biopsy images, it is useful to extract the rotation-invariant features by symmetrizing the co-occurrence matrix features, $\boldsymbol{M} + \boldsymbol{M}^\top$. The classifier weight $\boldsymbol{W}$ in constructing the discriminative co-clusters g_k (Sec.2.2) is thus symmetric as shown in Fig. 5. The EM method



$\Delta = 2$ $\Delta = 10$ $\Delta = 20$ $\Delta = 30$ Co-clusters g_k for $\Delta = 30$

Figure 5: Classifier weights for the discriminative co-clusters.

Method	disc. CoF+HoCoF	CoF [23]	[15]	[2]
EER (%)	**94.29**	93.03	91.60	89.29

Table 1: Comparison to the other methods.

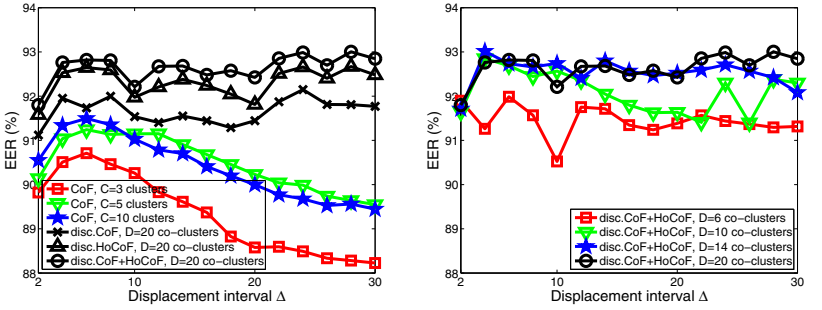
is slightly modified to cope with the symmetry[2] and to produce symmetric discriminative co-clusters g_k (Fig. 5).

For various intervals $\Delta \in \{2, \cdots, 30\}$, the performance results are shown in Fig. 6a: disc. CoF stands for the discriminative co-occurrence features using g_k in (2), disc. HoCoF for the higher-order co-occurrence features using $g_k g_l$ in (4), and disc. CoF+HoCoF indicates the concatenated features of the discriminative co-occurrence features and the higher-order co-occurrence ones, where we use $D = 20$ co-clusters. For comparison, we also applied the standard co-occurrence feature (1) [23] with f_i constructed by applying EM on the gray-scale pixel intensities, which is denoted by CoF with the number of clusters C. The performances of the standard CoF degrades as the interval Δ increases, while the proposed method produces stably high performances. As shown in Fig. 5, the co-occurrences of the near-by pixels are simply concentrated along the diagonal, while separated pixels form more complicated co-occurrences. The standard CoF based on $f_i f_j$ cannot capture such complicated co-occurrence characteristics of larger Δ, deteriorating the performance. The proposed method effectively exploit them by using the discriminative co-clusters. Note that the dimensionality of CoF with $C = 10$ clusters is 55 by considering symmetry, compared to only 20 in disc. CoF with $D = 20$ co-clusters.

We then investigate the effect of the number of discriminative co-clusters D on the performance of disc. CoF+HoCoF. The results are shown in Fig. 6b. In the case of only six discriminative co-clusters, the performances are slightly degraded since such co-clusters are too few to characterize the discriminative co-occurrences, but the performances are greatly improved by using sufficient number of co-clusters.

Finally, the performance of the proposed method is compared to those of the other methods; color coherent vectors [15] which is extended from color histogram, and the generic feature for cancer detection [2]. For the proposed method disc. CoF+HoCoF and the method CoF [23], all the features of various Δ's are concatenated into the (long) feature vector. The results are shown in Table 1. The proposed method produces superior performances to the others by effectively exploiting the discriminative co-occurrence characteristics. Fig. 7 shows the classification (detection) results by the proposed method in pseudo colors. The results are quite similar to the ground truth labels given by the expert pathologists, demonstrating that the proposed method can efficiently facilitate the inspection which has been exhaustively performed by hand.

[2] see supplementary material for details.



(a) Comparison to the standard CoF (b) Number of co-clusters D

Figure 6: Performance results on various displacement intervals Δ.

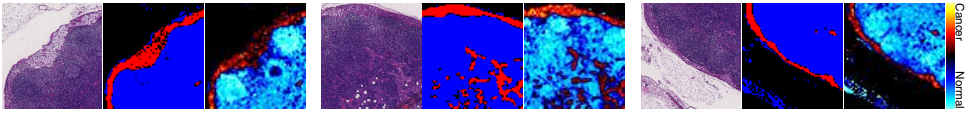


Figure 7: Detection results. In each block, left: biopsy image, middle: labels for cancer (red) and normal (blue), right: classification outputs by the proposed method in pseudo colors.

4.2 Pedestrian detection

Next, we evaluated the performances of the proposed method on pedestrian detection by using the Daimler Chrysler pedestrian benchmark dataset, created by Munder and Gavrila [14]. The dataset is split into five disjoint sets, three for training and two for test. Each set has 4,800 positive (pedestrian) and 5,000 negative (non-pedestrian) images of 18×36 pixels, as shown in Fig. 10. We measure the equal error rate according to the standard protocol for the dataset, training on two out of three training sets at a time and testing on each of the test sets.

We follow the approach to classify pedestrian images by utilizing orientations of image gradients [5, 8]. Namely, in this experiment, the primitive quantitative data are gradient orientations, $\theta = \arctan(\partial_x I, \partial_y I)$ where $\partial_x I, \partial_y I$ are derivatives of an image I along x- and y-axes, respectively. Such quantitative representation differs from the pixel intensities in that the orientation is cyclic in $\theta \in [0, 2\pi)$, and for constructing discriminative co-clusters (Sec.2.2), we slightly modify the EM method so as to cope with the cyclic orientations[3]; the obtained co-clusters g_k are shown in Fig. 8. The subsequent procedures in the proposed method, however, are the same as in the cancer detection (Sec.4.1) using the pixel intensities.

In a preprocessing, the derivatives are computed by applying Roberts filters to provide the gradient orientation θ and magnitude $w = \sqrt{\partial_x I^2 + \partial_y I^2}$ at each pixel. The gradient orientation is finely partitioned into 90 bins ($L = 90$ in Sec.2.2), and the gradient magnitude $w(\boldsymbol{p})$ is used for the weighting in the feature extraction; $\omega(\boldsymbol{p}, \boldsymbol{q}) = \min[w(\boldsymbol{p}), w(\boldsymbol{q})]$ in (2) and $\omega(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}, \boldsymbol{s}) = \min[w(\boldsymbol{p}), w(\boldsymbol{q}), w(\boldsymbol{r}), w(\boldsymbol{s})]$ in (4) for suppressing isolated noises [8]. In the pedestrian detection, pedestrians are usually upright in images and the object orientation could be a clue for classification. Therefore, the co-occurrence is computed by using various displacement directions, $\mathbb{N} = \{(\boldsymbol{p}, \boldsymbol{q}) | \boldsymbol{q} - \boldsymbol{p} = \boldsymbol{\Delta}\}$ where $\boldsymbol{\Delta} \in \{(2,0), (2,2), (0,2), (-2,2)\}$. The discriminative co-clusters are obtained for the co-occurrences of respective displacements, and they are asymmetric as shown in Fig. 8; note that they are cyclic along the both axes. In addition, as in [8, 11], the images (18×36 pixels) are spatially partitioned into 3×4 bins for extracting the parts-based features.

[3] see supplementary material for details.

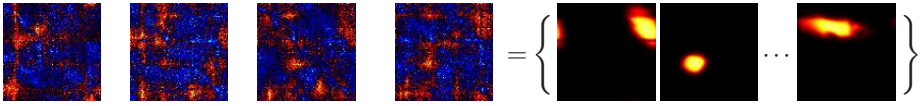


Figure 8: Classifier weights for the discriminative co-clusters.

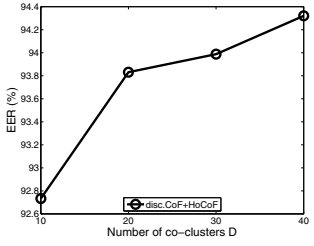
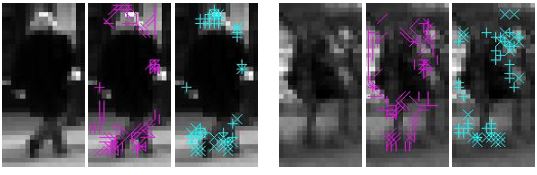


Figure 9: Performances on various number of co-clusters D.



(a) True positive image (b) False positive image

Figure 10: Pairs and quadruplets.

Method	disc. CoF+HoCoF	disc. CoF	CoF [8]	HOG [5]	[12]	[14]	[21]
EER (%)	**94.32**	93.80	93.68	86.41	89.25	89.82	91.10

Table 2: Comparison to the other methods.

Fig. 9 shows the performance of the proposed method disc. CoF+HoCoF by varying the numbers of co-clusters, $D = 10 \sim 40$. Smaller number of co-clusters $D = 10$ are not sufficient to exploit the characteristics of such complicated weights as shown in Fig. 8, while the performances are improved on larger numbers of co-clusters $D \geq 20$.

Then, we show the pairs (doublets) and quadruplets that highly contribute to positive classification for pedestrian in Fig. 10; the pairs are indicated by the (magenta) lines and the quadruplets are by the (cyan) cross lines. In the pedestrian image, the primary pairs and quadruplets are mainly located around the head, shoulder and foots (Fig. 10a). This shows that the proposed method extracts pedestrian-specific characteristics. The non-pedestrian image that is misclassified has also such pairs and quadruplets on the parts which are similar to the foots and shoulder of the pedestrian (Fig. 10b).

The performance of the proposed method is compared to those of the other methods in Table 2; for comparison, we applied the standard co-occurrence method [8] which predefines f_i of bilinear interpolated nine orientation bins, and HOG features [5], and in the other methods [12, 14, 21], we show the performances reported in the respective reference papers. We used disc. CoF+HoCoF with 40 co-clusters ($D = 40$), which is half number of clusters used in the standard CoF [8] with nine orientation bins producing 81 types of $f_i f_j$. The proposed method produces favorable performances compared to the others, demonstrating that the discriminative and higher-order co-occurrences are also effective to the quantitative data of (cyclic) image gradient orientations.

5 Conclusion

We have proposed the method to extract image features based on effective higher-order co-occurrences. The proposed method first constructs the discriminative co-clusters to directly quantize pair-wise quantitative data in the joint space, whereas the standard methods uti-

lize simple clusters produced in an unsupervised manner for quantizing point-wise data. By utilizing the discriminative co-clusters, the co-occurrence characteristics are effectively extracted in a discriminative way with a fewer number of cluster components, *i.e.*, low dimensional features. Therefore, the co-occurrences can be extended to the higher-order co-occurrences of feasible dimensionality by considering quadruplets, pairs of pair-wise data represented efficiently by the discriminative co-clusters. The higher-order co-occurrence features take into account the relationships in the multiplets more than only doublets (pairs) to capture richer information of image textures. In the experiments on image classifications for cancer cells and pedestrians, the proposed method exhibited favorable performances compared to the other methods, even to the standard co-occurrence based methods.

Our future works include to automatically determine the number of co-clusters such as by introducing AIC in the EM method and to apply the method to multi-class problems.

References

[1] Y-L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[2] A. Brook, R. El-Yaniv, E. Isler, R. Kimmel, R. Meir, and D. Peleg. Breast cancer diagnosis from biopsy images using generic features and svms. Technical Report CS-2008-07, Technion - Israel Institute of Technology, 2008.

[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[4] G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *ECCV Workshop on Statistical Learning in Computer Vision*, pages 59–74, 2004.

[5] 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.

[6] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-3(6):610–621, 1973.

[7] J. Huang, S. R. Kumar, M. Mitra, W-J. Zhu, and R. Zabih. Image indexing using color correlograms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–768, 1997.

[8] T. Kobayashi and N. Otsu. Image feature extraction using gradient local auto-correlations. In *European Conference on Computer Vision*, pages 346–358, 2008.

[9] T. Kobayashi and N. Otsu. Bag of hierarchical co-occurrence features for image classification. In *International Conference on Pattern Recognition*, pages 3882–3885, 2010.

[10] H. Ling and S. Soatto. Proximity distribution kernels for geometric context in category recognition. In *International Conference on Computer Vision*, pages 1–8, 2007.

[11] D.G. Lowe. Distinctive image features from scale invariant features. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[12] S. Maji and A.C. Berg. Max-margin additive classifiers for detection. In *International Conference on Computer Vision*, pages 40–47, 2009.

[13] N. Morioka and S. Satoh. Building compact local pairwise codebook with joint feature space clustering. In *European Conference on Computer Vision*, pages 692–705, 2010.

[14] S. Munder and D. M. Gavrila. An experimental study on pedestrian classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1863–1868, 2006.

[15] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *ACM International Conference on Multimedia*, pages 65–73, 1997.

[16] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[17] R. Rautkorpi and J. Iivarinen. A novel shape feature for image classification and retrieval. In *International Conference on Image Analysis and Recognition*, pages 753–760, 2004.

[18] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[19] V.N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[20] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *European Conference on Computer Vision*, pages 705–718, 2008.

[21] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[22] Y. Xiao and G. Xuan. Fast em algorithm of multi-dimensional histogram in medical images. In *International Conference on Diagnostic Imaging and Analysis*, pages 328–333, 2002.

[23] A. Yaguchi, T. Kobayashi, K. Watanabe, K. Iwata, T. Hosaka, and N. Otsu. Cancer detection from biopsy images using probabilistic and discriminative features. In *International Conference on Image Processing*, pages 1641–1644, 2011.

[24] Y. Yang and S. Newsam. Spatial pyramid co-occurrence for image classification. In *International Conference on Computer Vision*, pages 1465–1472, 2011.