Sample Space Dimensionality Refinement for Symmetrical Object Detection

Yun-Fu Liu, *Member, IEEE*, Jing-Ming Guo, *Senior Member, IEEE*, Chih-Hsien Hsia, *Member, IEEE*, Sheng-Yao Su, and Hua Lee, *Fellow, IEEE*

Abstract—Formerly, dimensionality reduction techniques are effective ways for extracting statistical significance of features from their original dimensions. However, the dimensionality reduction also induces an additional complexity burden which may encumber the real efficiency. In this paper, a technique is proposed for the reduction of the dimension of samples rather than the features in the former schemes, and it is able to additionally reduce the computational complexity of the applied systems during the reduction process. This method effectively reduces the redundancies of a sample, in particular for those objects which possess partially symmetric property, such as human face, pedestrian, and license plate. As demonstrated in the experiments, based upon the premises of faster speeds in training and detection by a factor of 4.06 and 1.24, respectively, similar accuracies to the ones without considering the proposed method are achieved. The performance verifies that the proposed technique can offer competitive practical values in pattern recognition related fields.

Index Terms—Sample refinement, dimension reduction, data reduction, face detection, pedestrian detection.

I. INTRODUCTION

IN RECENT years, the continuing progress of the computing efficiency and parallel processing have triggered some of the subareas in computer vision, i.e., pattern recognition, continually overcoming the former bottlenecks on processing efficiency and accuracy related performance. For instance, the pedestrian detection has been mounted on some of mass-produced vehicles, and the current self-propelled robot is able to detect the moving path of any object during stable walking for avoiding collision [1]. Yet, even though the computational performance is continuously evolving, the associated development of high complexity algorithms still makes real-time processing infeasible.
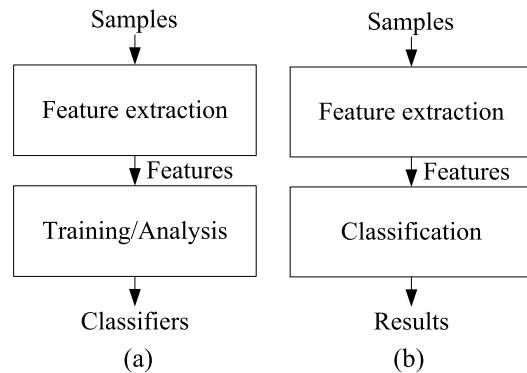
Fig. 1. Simplified conceptual (a) training and (b) classification flowcharts of a general pattern recognition system.

In general, the training and classification flowcharts of a general pattern recognition system can be illustrated in Fig. 1. First, the samples of interest, e.g., human face or pedestrian, should be captured for being described (feature extraction). Normally, some preprocessing/calibration methods are adopted to cope with the variations of capturing conditions such as illumination, contrast, or the mounted position/angle before the feature extraction [2], to yield a more stable feature distribution. During the collection of the samples in the training phase, a great number of samples is preferred to completely simulate the characteristics of the object of interest for a higher reliability of the trained classifier(s). This suggests that the increase of the number of samples will lead to exponentially increasing in training time to maintain a similar sample rate in hyperdimension space (caused by the curse of dimensionality [3]). Based upon this premise, the core of reducing the computational complexity of the entire system in terms of training or classification, should be around the feature extraction because of its complexity is tied up with the increased number of samples. Some available ways to reach this can be low-computational-required features [4], [5], higher efficient training/classification [6]–[10], or parallelism [11], yet these topics have been widely-discussed in the past decades. Thus, in this study we try to explore the additional possible room for the further reduction of the required computations of an entire algorithm. The induced benefit from this study can be further applied for a low-level hardware, or enabling an algorithm with a higher computation complexity since it can be reduced with the proposed scheme.

From another perspective that exploiting fewer but more representative features for training/classification can achieve the aforementioned goal, in fact the existing dimensionality
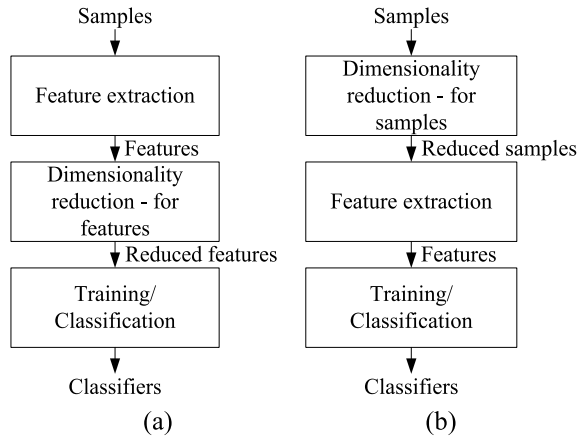
Fig. 2. Simplified conceptual training/classification flowchart with considerations of Dimensionality Reduction (DR) for (a) the extracted features and for (b) the captured samples.

reduction (DR) methods have confirmed this claim. Such as the former principal component analysis (PCA) [12], it can yield low-dimension data through ignoring the projected information onto those less important dimensions (smaller eigenvalues); the independent component analysis (ICA) [13] requires the distribution of the collected samples as non-Gaussian while maximizing the independence of each other during dimension reduction. Figure 2(a) demonstrates a normal flowchart geared with the DR. As it can be seen, the DR is usually employed after the feature extraction, thus although it reduces the time during training/classification, the influence of the increased number of samples is still not completely coped with, because the features of all samples still have to be calculated. An alternative thought as shown in Fig. 2(b) which simply exchanges the order of the feature extraction and the DR, the benefit of the dimension reduction can be enjoyed by not only the reduced data of the collected samples, but also the required computed features during training/classification. This type of method which reduces the dimension of samples, and performs at another timing point, is termed sample refinement (SR) in this paper.

In this paper, a general SR concept is highlighted and discussed for coping with any types of samples. As a result, both of the training and classification time can be reduced, and the corresponding descriptions are detailed in subsection 2.1. To solve the former iterative-based schemes which may reduce the overall processing efficiency, a simplified but of high practical value SR method, namely specifying sample space dimensionality refinement (S^3DR), is proposed to achieve the above demands as long as the samples are of partially symmetric property. This part is discussed and analyzed in subsection 2.2. In subsection 2.3, a well-known face detection method is adopted as a practical instance for the explanation of the corresponding alternatives and considerations using this proposed concept via an implantation. In Section 3, the proposed method is evaluated and compared to support its perspective and generalization capability. According to these results, the concept of SR is demonstrated, and it works well to acquire both benefits of dimensionality reduction and complexity reduction with three famous benchmarks, face,

pedestrian, and license plate. Finally, some conclusions of this study are drawn in Section 4.

II. SPECIFYING SAMPLE SPACE DIMENSIONALITY REFINEMENT

In general, in a determined dimensional space, e.g., the space constructed with multiple features as used in pattern recognition, whether the collected samples possess statistical significance determines if they are sufficient to completely express the characteristics of an object of interest. Normally, more samples are requires to yield a higher capability of fully expressing the object of interest, yet a linearly increase on computational complexity is its deficiency because of the corresponding feature calculations and subsequent training/classification as indicated in Fig. 2. In another perspective, except for the increase of samples, the dimension of a space also affects the possible capability to enable a fully representation (as demonstrated below). However, reducing the dimensions of the feature space also suggests employing fewer features to describe the collected samples. Consequently, the possibility that a feature distribution of a specific class overlaps to other classes will be increased, and thus sacrifices the correct classification rate theoretically.

In this section, a thought toward local sample similarity is brought out. It is able to utilize a fewer sample data to reach a more complete expression for an object of interest, meaning less computations is required. In addition, the accuracy decreases very limited. The following descriptions are separated into three parts, including theory, simplification, and application. Herein, the theory provides a general way for covering any types of samples; the simplification reduces the required computations while maintaining the property of the collected samples; the application provides a practical case to demonstrate the implementation and realization of the proposed concept.

A. Theory and a General Method

Normally, obtaining a classifier by involving the hard rules [14], e.g. decision tree, the targeted analysis and description is necessary as a crucial step. For those image samples of a specific object, some external components, e.g., color, luminance, or contour, depict its characteristics. Theoretically, when these components are of high variation, more information of this object can be offered. However, the similarity among each signal/feature, as termed redundancy in this paper, degrades the average of the supported information and the fully expression capability of the object. For instance, a sample set (S) defined as

$$S = \{s_p | p = [1, P], s_p \in \mathbb{Z}^{M \times N}\}, \quad (1)$$
$$s_p = \{x_t^p | t = [1, M \times N], x_t^p = [x_{min}, x_{max}]\} \quad (2)$$

in which P denotes the number of collected samples of dimension $M \times N$; x_{min} and x_{max} denote the theoretical lower and upper bounds of the possible values of x_t^p. Subsequently, the ability of the fully expression for objects $\propto \frac{P}{\|s_p\|}$, which not only affected by the number of samples, but also the
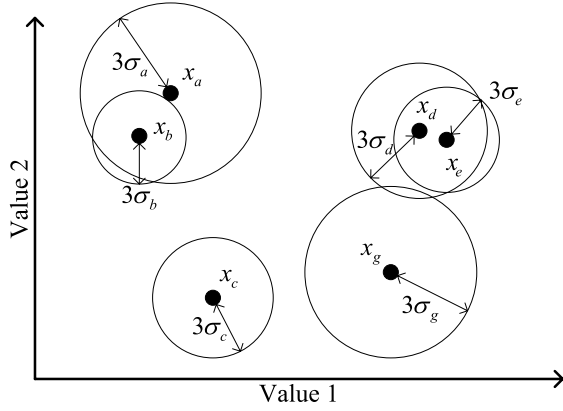
Fig. 3. Conceptual diagram of the projected sample distributions on a representative value plane.

space dimension where samples are represented. Moreover, the former additionally implies the difficulty of sample collection, and both of them suggest the consumed time during the training phase from the aspect of carried data amount (i.e., in the Viola-Jones's AdaBoost [15], when training with the sample of sizes 24 × 24 or 32 × 32, the corresponding number of weak classifiers are 162,336 and 510,464, respectively. Thus, the required training time is increased proportionally by a factor of 3.14 (=510464/162336; same issue is also discussed in Section 3)).

To maximize the entire information, the signal/feature with less significance or even totally replaceable significance (redundancy) can be ignored or further analyzed. Thus, when the samples can be represented with fewer data, fewer computations can be achieved during the subsequent feature extraction and training/classification as illustrated in Fig. 2. Theoretically, the redundancy happens when multiple constructed information x_t^p are similar to each other, which can be modeled with the restriction as formulated below,

$$\forall p: \left\| x_a^p - x_b^p \right\|_2 \leq \gamma, \tag{3}$$

where $a, b = [1, M \times N]$. In which, all of the collected samples labeled with p should conform to the same restriction as indicated. Yet, another perspective, if consider the x_t^p's group of the same property as an identical x_t^p (e.g., same value or described feature), the dimension of sample space $\|s_p\|$ can be reduced with a bounded γ, while gearing with a similar capability of expressing the object of interest. Figure 3 demonstrates an instance of projected sample's distribution on a representative plane, where each circle denotes the distribution of a specific x_t^p of all samples p, and $3\sigma_t$ denotes the ranges of these distributions because the Gaussian distribution is supposed for modeling. As it can be seen, both pairs (x_a, x_b) and (x_d, x_e) are reasonably merged with a predetermined γ to independent x_ts, and thus the objective of the SR can be achieved.

For reaching this purpose, a theoretical sample refinement algorithm based upon the thought of Eq. 3 is organized as below for the general cases. In addition, the corresponding diagrams are drawn in Fig. 4(a) for a better understanding. First, a preprocessing stage for the collected sample set S as
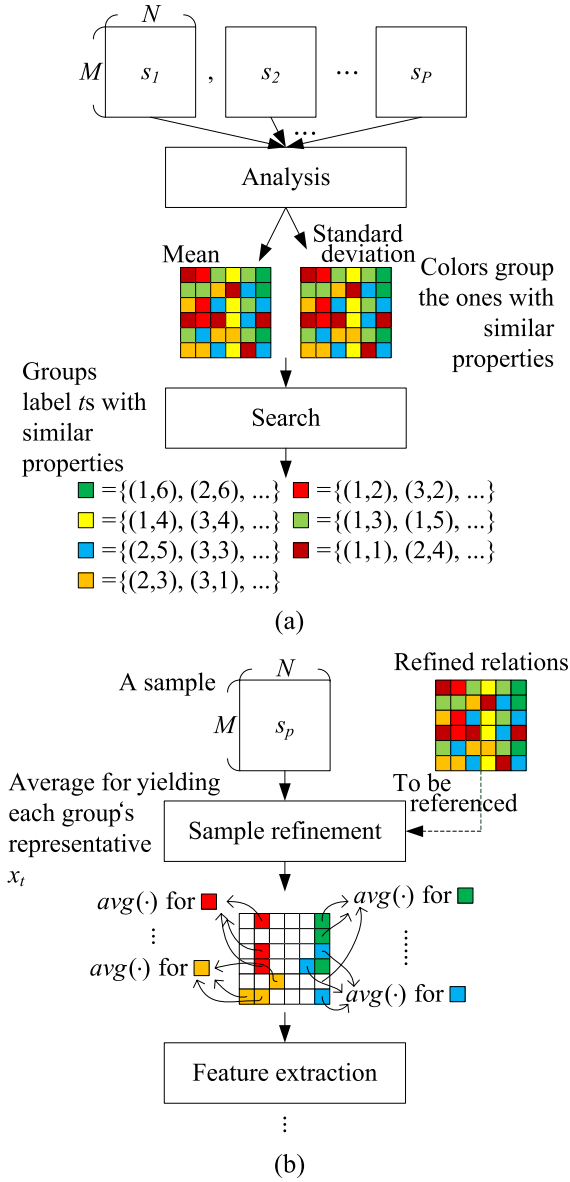

Fig. 4. Sample refinement conceptual algorithms for general cases. (a) Samples analysis. (b) Practical application.

defined in Eq. 1 should be performed along with the following steps:

1) Calculate the mean and standard deviation of all samples in S for the further processing. The corresponding formulations are shown below:

$$\mu_t = \frac{1}{P} \times \sum_{i=1}^{P} x_t^i, \tag{4}$$

$$\sigma_t = \frac{1}{P} \times \sqrt{\sum_{i=1}^{P} \left(x_t^i - \mu_t \right)^2} \tag{5}$$

where the definitions of P, x_t^i, and t are identical to those defined in Eqs. 1 and 2.

2) Initialize the $\mu'_g = \mu_1$, $\sigma'_g = \sigma_1$, and $Q = 1$, where $g = [1, Q]$ and Q denotes the number of groups (containing similar $x_t(s)$; i.e., the number of pixels with different colors as shown in Fig. 4, which is seven in this case).

3) Compare the pairs $\left(\mu'_g, \sigma'_g\right)$ and (μ_t, σ_t) with the following restrictions:

$$\left\| \mu_t - \mu'_g \right\|_2 \le \gamma_\mu, \tag{6}$$

and meanwhile it conforms to

$$\left\| \sigma_t - \sigma'_g \right\|_2 \le \gamma_\sigma, \tag{7}$$

where $t = [2, M \times N]$. If true, add t into $G_g = \{\dots, t\}$, and recalculate the μ'_g and σ'_g with all of x_ts in gth group for updating; if false, add t into $G_{Q+1} = \{t\}$, and update $Q = Q+1$, while $\mu'_{Q+1} = \mu_t$ and $\sigma'_{Q+1} = \sigma_t$ for the subsequent merging. Notably, the g should be within $[1, Q]$, and the next pair of $(\mu_{t+1}, \sigma_{t+1})$ is estimated after the current estimation.

4) If any grouping happens during the previous merging step, repeat Step 3 until no grouping operation happens; if not, exports the arranged Q relations $G_1, \dots, KG_Q$ for the sample refinement.

With this analysis, those x_ts of high similarity which associate to a specific group are known, while the extracted relations can be utilized in any training/classification procedure, since it simply needs the above independent preprocessing on samples. Figure 4(b) demonstrates a corresponding example. First, when any collected sample inputs into the training/classification system, the sample should be refined with the extracted relations for complying the SR. As it can be seen in the conceptual colored relation located at the top-right corner, all of the x_ts mapping to an identical color of the same relation are considered in the same group. In this case, we simply take the green, red, blue and yellow cases for demonstration, and the rest white pixels can be processed likewise. In practice, the statistical mean of those x_ts of the group G_g is regarded as a new value to represent the sample for further processing.

B. Simplification for Samples With Symmetric Property

In the field of pattern recognition, no matter what the object of interest is, the digital signals are in two-dimension (2D) with spatial relation. Meanwhile, the number of the pixels indicates the size of data. Since these pixels normally possess spatial relationship rather than independent to each other, thus apart from the ways using statistical characteristics as the features for description, some others considered that the spatial contexts have to be exploited in pattern recognition, e.g., the edge and orientation [16]. This suggests that the spatiality of a sample has to be retained, even after the process of SR, for fitting the extracted features which were designed for this specific 2D digital signal. Under this consideration, the given concept of the SR as detailed in the previous subsection should be further adjusted accordingly, and thus an extended SR method is proposed, namely specifying sample space dimensionality refinement (S^3DR). This method is able to maintain the spatial nature of a 2D sample as well as require only simple and few calculations (without iteration strategy in the previous general method) to obtain the benefit of the SR, simultaneously.

In simple terms, the discussed general SR method collects x_t with similar statistical properties as the represented colors in the Fig. 4(a), yet after the process of "search," no relationship exists among groups. From another viewpoint, the samples with symmetric property, such as human face, pedestrian, and license plate, are popular issues in pattern recognition; high similarity between its left and right parts of the axis of symmetry, the two parts can be considered as connected groups for their high similarity as indicated in previous subsection. As shown in Fig. 5, samples from the three different types of objects are adopted for demonstrating this observation. Herein, the foreground (containing the object of interest) and background (without object) each possesses 1000 samples, in which the MIT CBCL database [17] is employed for human face; the INRIA database [18] is for pedestrian, and the NTUST MSPLab database [19] is for license plate. In each subfigure, the indicated points denote the sampled x_t^ps which are spatially symmetric from the red axis of symmetric, and the corresponding standard deviation distributions are derived as below,

$$h_r = \sum_{p=1}^{P} \begin{cases} \dfrac{1}{P}, & \text{if } \left[0.5 \times \sqrt{\sum_{i=1}^{2} \left\{ x_i^p - \dfrac{(x_1^p - x_2^p)}{2} \right\}^2} \right] = r, \\ 0, & \text{O.W.} \end{cases}$$
$$\text{where } r \in \mathbb{Z} \tag{8}$$

where $\left(x_1^p, x_2^p \right)$ denotes a symmetric pair; $[\cdot]$ denotes round off operation; P denote the number of samples; r denotes the scale of the horizontal axes of the curves in Fig. 5. As it can been seen, the uniform distribution of the background x_t^p relatively reflects the corresponding h_r's distribution similar to a Gaussian shape. Yet, it is obvious that the h_r distributions of the foregrounds are still more stable than that of the background because of the mentioned symmetric property of these types of samples. This indicates that even this symmetric property is adopted for dimension reduction of these samples, the fully expression capability is simply somewhat reduced. Consequently, for the samples with symmetric property, the following equation can be applied before the feature extraction as shown in Fig. 2(b). This can meet the objective of the SR in sample reduction, while simply inducing limited sample description distortion. First, the s_p as shown in Eq. 1 is redefined to a 2D signal as $s_p = \{x_{i,j}^p | i = [1, M], j = [1, N]\}$, and thus the SR samples can be generated with

$$s'_p = \left\{ x'_{i,j} | x'_{i,j} = \frac{x_{i,j} + x_{i,M-j}}{2}, i = [1, M], j = [1, \lfloor N/2 \rfloor] \right\},$$
$$\text{and} \quad s'_p \in \mathbb{R}^{M \times \frac{N}{2}} \tag{9}$$

where $\lfloor \cdot \rfloor$ denotes the round down operation; M and N are defined in Fig. 4. Notably, all of the samples with detectable symmetric property are able to be applied with this concept, while the definition of a corresponding affined calculation of Eq. 8 is needed.

C. A Practical Instance and Related Analysis

In this subsection, a concrete case of employing the proposed S^3DR is presented. Moreover, for highlighting the
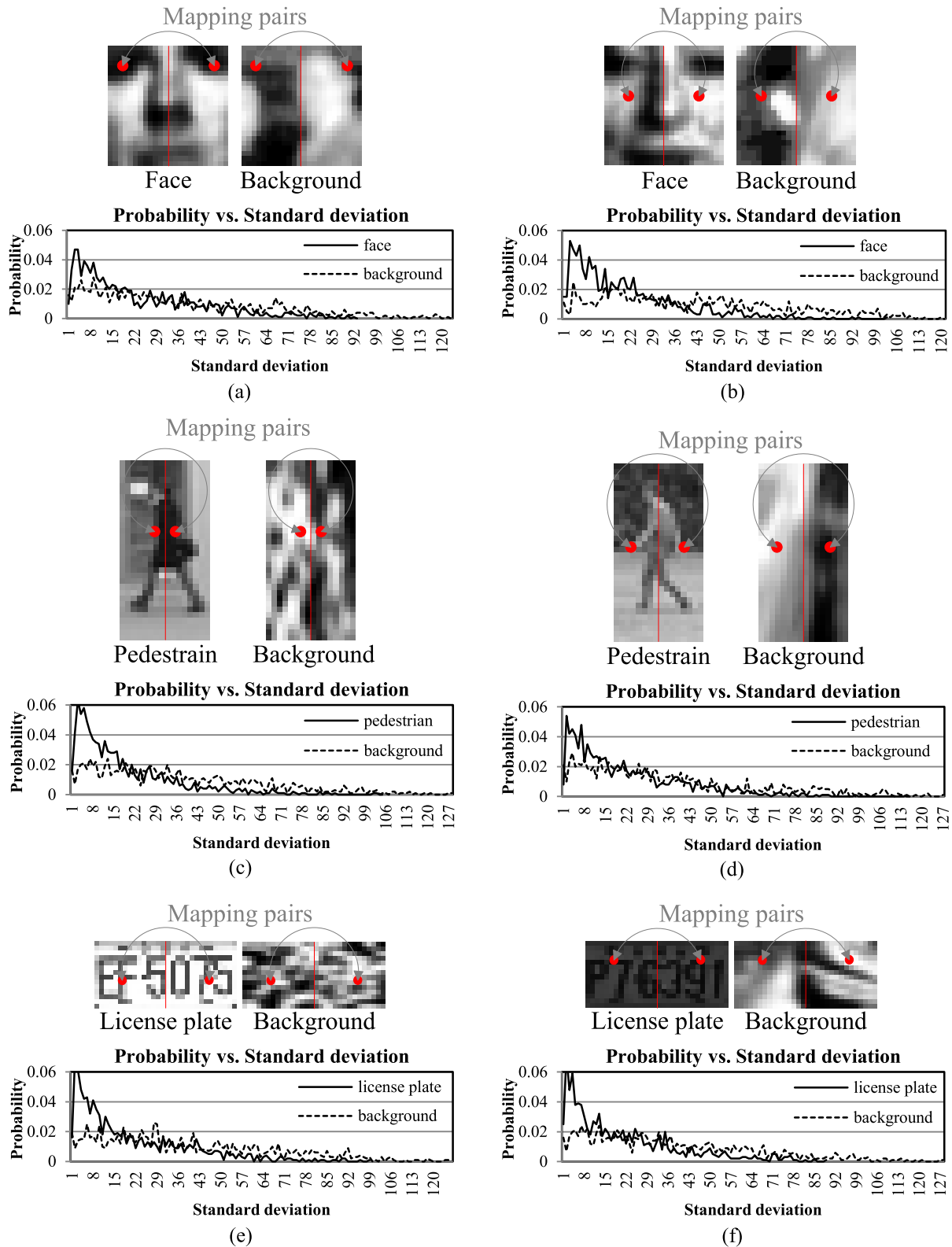
Fig. 5. Examples of distribution stability of the indicated symmetric pair $\left(x_1^p, x_2^p\right)$ on various types of symmetric samples, including (a)-(b) human face [17], (c)-(d) pedestrian [18], and (e)-(f) license plate [19]. Herein, the red line denotes the supposed axis of symmetric, and labeled points denote where we sampled.

symmetric properties we observed, the well-known Viola-Jones's AdaBoost face detection [15] is adopted as the infrastructure to be modified. The difference of the corresponding performance through comparing the traditional and the modified ones reflects the characteristics of S^3DR (the results are shown in the next section). In the following,
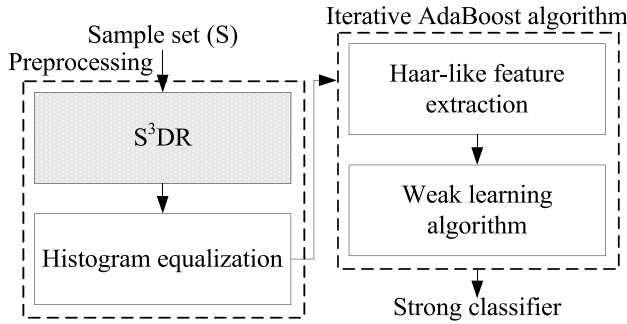
Fig. 6. Simplified AdaBoost training algorithm [15] with S³DR. (The different part comparing with the traditional algorithm is colored with gray.)

Fig. 8. Samples of Haar-like features [15]. (a)-(e) Feature types 1-5, respectively, from left to right.

TABLE I

COMPARISON OF NUMBER OF WEAK CLASSIFIERS BETWEEN TWO DIFFERENT SAMPLE SIZES 24×24 AND 24×12

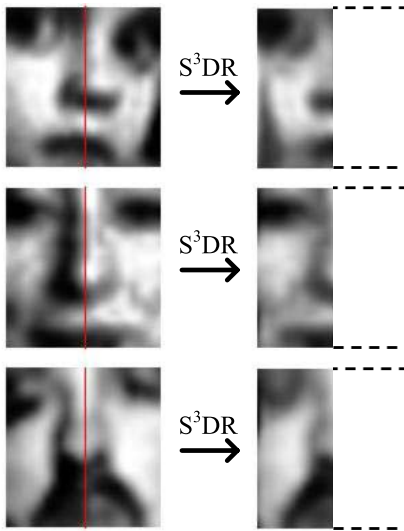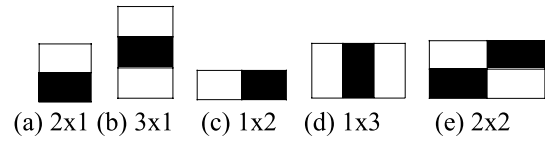| Types of Haar-like feature (definitions in Fig. 8 is used; each mxn is provided below)	Comparison of numbers of weak classifiers (Eq. 10 is adopted for simulation)
Type 1 (2x1)	$\dfrac{R_1(24,12)}{R_1(24,24)} = \dfrac{11232}{43200} = 0.26$
Type 2 (3x1)	$\dfrac{R_2(24,12)}{R_2(24,24)} = \dfrac{7176}{27600} = 0.26$
Type 3 (1x2)	$\dfrac{R_3(24,12)}{R_3(24,24)} = \dfrac{10800}{43200} = 0.25$
Type 4 (1x3)	$\dfrac{R_4(24,12)}{R_4(24,24)} = \dfrac{6600}{27600} \cong 0.24$
Type 5 (2x2)	$\dfrac{R_5(24,12)}{R_5(24,24)} = \dfrac{5184}{20736} = 0.25$

Fig. 7. Face samples refined by S³DR.

the descriptions are separated into training and detection two phases, while only the altered parts and related analysis are discussed. The details of the AdaBoost please kindly further refer to the former article [15].

1) Training Phase: Figure 6 illustrates a simplified algorithm of the AdaBoost training system which can be separated into two parts as preprocessing and its iterative algorithm. First, when a dataset (S) is imported, dimension of each sample (human face in this instance) should be reduced through Eq. 9 for SR. Figure 7 shows some results for a better understanding. Obviously, even in the case that the sample is not fully symmetric as the top case, the corresponding SR result still stably behaves for further analysis. Subsequently, the histogram equalization [20] is performed on the sample of half size for the stabilization of the sample luminance. These preprocessed samples are then processed with the AdaBoost's weak learning algorithm which includes extracting the representative Haar-like features (the five basic types are shown in Fig. 8) as well as the subsequent strong classifier construction.

The influences induced from S³DR are discussed as follows. According to Lienhart et al.'s formula [21] as defined below, the number of the possible weak classifiers using Haar-like

feature can be derived,

$$R_{type}(M, N) = X \times Y \times \left(N + 1 - n\frac{X+1}{2}\right)$$
$$\times \left(M + 1 - m\frac{Y+1}{2}\right) \quad (10)$$

where $X = \frac{N}{n}$ and $Y = \frac{M}{m}$; $M \times N$ is defined in Fig. 4; m and n denotes the smallest size of the *type*th Haar-like feature as shown in Fig. 8; X and Y denote the greatest scales the *type*th feature can be enlarged. Apparently, the sample size is directly proportional to the number of weak classifiers. On the other hand, since this number also affects the training time during the weak learning algorithm, the ratio between before and after SR process can be calculated for the theoretical estimation. The corresponding ratios are organized in Table I which involves five types of basic Haar-like features for a fair comparison. As it can be seen in the corresponding results, around 0.24 to 0.26 proportions of the number of weak classifiers to the original sample size 24×24 can be achieved after the SR process. Since the complexity of the weak learning algorithm relies on the pool size of the weak classifier candidates, the learning time is decreased linearly as expected. Moreover, since the weak classifiers are derived from the SRed samples, the descriptive capability of each weak classifier is improved as well, and thus using fewer classifiers can achieve equivalent performance as expected. The corresponding experiments are shown and discussed in next section.

2) Detection Phase: Figure 9 shows a rough Viola-Jones' face detection algorithm [15], in which the modified part by embedding the S³DR is colored with gray for comparison. First, for the extracted Haar-like features as that extracted during the training phase, the S³DR is performed for the
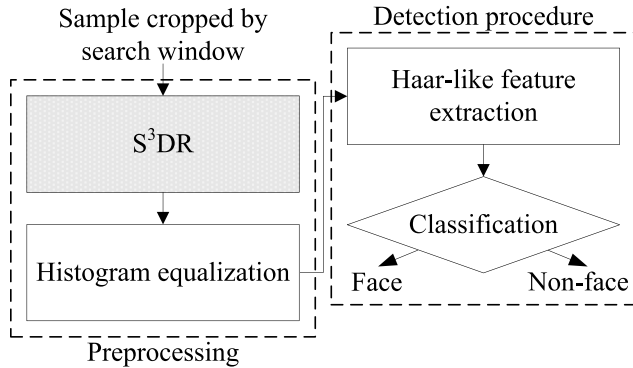
Fig. 9. Simplified AdaBoost face detection algorithm [15] geared with the S³DR (the different part compared to the traditional algorithm is colored with gray).

sample refinement. Subsequently, since the dimension of the sample is reduced, only half of the original samples are then enhanced through the histogram equalization, and thus less computation can be achieved. Comparing with the traditional AdaBoost face detection algorithm, simply 50% of the related computations are needed because of the difference of the enhanced areas.

III. EXPERIMENTAL RESULTS

In this section, the performance of the proposed S³DR is evaluated with two aspects, time consumption and detection accuracy. In the following experiments, Viola-Jones AdaBoost algorithm as introduced in subsection 2.3, is considered as the baseline for comparison. Moreover, the MIT CBCL human face dataset [17] is also adopted, in which 2,429 faces and 4,548 non-faces are contained for the classifier training, while another set of 472 faces and 23,573 non-faces are employed for test. Notably, all of the original samples are bilinear interpolated to 24 × 24 for evaluation. Herein, the test platform is with Intel i3-2120 dual-core processor and 4GB RAM.

Table II exhibits the comparison in terms of time consumption using the AdaBoost algorithm with and without S³DR, in which the influence induced from this SR is the reduced number of weak classifiers. Notably, the "window" (processed windows per second) mentioned in this table is of size 24 × 24 as defined above for consistency. Obviously, training time is reduced accordingly by a factor of 4.06 (12668.183/3117). According to the theoretical analysis as discussed in subsection 2.3.1, the expected reduction in time consumption is obtained (0.246 = 1/4.06 which falls into the range in between 0.24 and 0.26 as we demonstrated). Moreover, the reduction scale depends on the target object of interest (e.g., human face, pedestrian, or license plate), since they affect the utilized numbers and proportions of the five basic types of Haar-like features as shown in Fig. 8. In addition, the processing efficiency by considering S³DR can yield a performance about 1.24× (=1615.51/1303.78) faster than the traditional one without this consideration. As discussed in subsection 2.3.2, even though S³DR requires additional computations as formulated in Eq. 9, due to its simple calculations, the benefit obtained from that is only half of the samples required to be enhanced through the

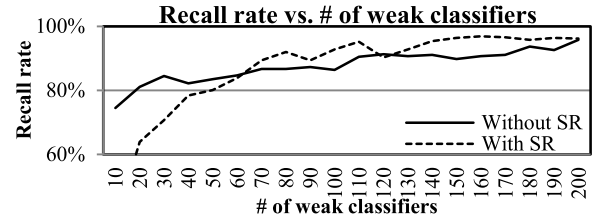| Stages \ SR	Without SR	With SR
Training time (minutes)	12,668	3,117
Processed windows per second	1,304	1,616

Fig. 10. Comparison on detection rates of various numbers of weak classifiers. In which, the thresholds used in with and without SR are 0.46 and 0.49, respectively.

histogram equalization. Thus, in overall a faster processing speed can be achieved. In summary, due to the proposed method reduces redundancies only for samples rather than classifiers, it can be freely adapted to other learning, detection, or classification algorithms to gain the same benefit of complexity reduction.

Figure 10 shows the result between recall rate (*Rec*) as defined below and the number of weak classifiers to see the saturation condition for computation complexity,

$$Rec = \frac{tp}{tp + fn}, \qquad (11)$$

where *tp* and *fn* denote the numbers of true positives and false negatives, respectively. To provide a fair comparison, the thresholds of yielding similar recall rate with and without SRs are determined: 1) Without SR: threshold = 0.49, recall rate = 95.75%, and false positive rate = 0.339%; 2) with SR: threshold = 0.46, recall rate = 96.178%, and false positive rate = 1.642%. This result demonstrates that the weak classifiers trained by the algorithm with S³DR have much representative capability than that without SR. For instance, according to this result, for achieving the 95% of recall rate, only 110 weak classifiers are required when S³DR is applied; conversely, the traditional detection method needs 198 weak classifiers to achieve the same performance. This verifies the statement addressed in subsection 2.3.2, and meanwhile this suggests that when a specific performance is demanded, a faster processing speed can be obtained because of the computations of fewer classifiers, which also has a linear relationship to the reduced numbers of weak classifiers.

To estimate the accuracy of the systems with and without the proposed S³DR, two benchmarks, including the previous MIT CBCL face dataset [17] and the pedestrian dataset, termed INRIA [18], are employed. In the former INRIA dataset, totally 2,416 and 12,180 samples of size 64 × 128 for pedestrian and non-pedestrian are conducted for training; during testing, 1,126 pedestrian samples of size 64 × 128

TABLE III
PERFORMANCE COMPARISON AMONG VARIOUS DATASETS AND METHODS WITH AND WITHOUT THE PROPOSED S^{3}DR

Methods / Datasets	INRIA pedestrian dataset [18]		MIT CBCL face dataset [17]	
	Without SR	With SR	Without SR	With SR
Viola-Jones [15]	0.81	0.82	0.98	0.97
Zhang et al. [22]	0.92	0.92	0.95	0.92
Abdullah et al. [23]	0.94	0.93	0.96	0.95

and 453 landscape photographs of any sizes are involved. Moreover, three methods including Viola-Jones's method [15], Zhang et al.'s method [22], and Abdullah et al.'s method [23], are all considered for comparison. In addition, to fully show the performances, the best F$_1$-score of the precision-recall curve is utilized for presentation as formulated below,

$$F_1 = 2 \times \frac{Pre \times Rec}{Pre + Rec}, \tag{12}$$

$$Pre = \frac{tp}{tp + fp}, \tag{13}$$

where *Rec* is defined in Eq. 10, and both *tp* and *fp* denote the true positive and false positive, respectively. Table III shows the corresponding results. As it can be seen, normally the performances are quite similar in most cases, yet the case, MIT CBCL [17] with Zhang et al.'s method [22], reveals about 3% degradation caused by the usage of the SR. According to our observation, when a feature provides less descriptive capability to their targeted objects, such as Zhang et al.'s local binary pattern (LBP) to face or Viola-Jones's Haar-like feature to pedestrian, unstable performances can yield, and thus some fluctuations may happen. Based upon these results, since the performances with and without the SR process are rather similar on average, the sample dimension can be safely reduced by the proposed S^{3}DR method to yield less computations as shown in Table II without apparently sacrificing the accuracy as demonstrated.

IV. CONCLUSIONS

In this paper, a sample refinement technique, termed S^{3}DR, is proposed to reduce the dimensions of samples for providing two benefits, dimensionality reduction and less computational complexity, during training and classification phases. Comparing with the former methods such as PCA, a great superiority is possessed, since the former methods require additional computations for the process of dimensionality reduction, while the proposed method can yield less computations to the entire systems, e.g., object detection. Notably, any future detection systems can integrate the proposed method for a higher processing efficiency. Fundamentally, the proposed S^{3}DR simplifies the procedure by using simple calculations when the target samples possess symmetric property. In fact, many targets in ongoing topics of pattern recognition meet this condition, such as human face, pedestrian, and license plate. As documented in experiments, higher processing efficiency on both training and classification phases are achieved by

maintaining a similar accuracy to the case without involving the SR. Moreover, the higher representative ability per weak classifier is endorsed as well, and all of the samples geared with the symmetric property are able to incorporate with the S^{3}DR theoretically as we discussed in subsection 2.2. Consequently, this is a novel research topic for further exploring and discovering, and lower accuracy degradation by applying SR process in the future work.

REFERENCES

[1] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. Robot. Res.*, vol. 21, no. 8, pp. 735–758, Aug. 2002.

[2] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1519–1524, Sep. 2006.

[3] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *Proc. 13th Annu. ACM Symp. Theory Comput.*, 1998, pp. 604–613.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[5] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, 2007.

[6] S. Kollias and D. Anastassiou, "An adaptive least squares algorithm for the efficient training of artificial neural networks," *IEEE Trans. Circuits Syst.*, vol. 36, no. 8, pp. 1092–1101, Aug. 1989.

[7] C. Charalambous, "Conjugate gradient algorithm for efficient training of artificial neural networks," *IEE Proc. G, Circuits, Devices, Syst.*, vol. 139, no. 3, pp. 301–310, Jun. 1992.

[8] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *J. Mach. Learn. Res.*, vol. 4, no. 6, pp. 933–969, Dec. 2004.

[9] P. Sharma, K. V. Arya, and R. N. Yadav, "Efficient face recognition using wavelet-based generalized neural network," *Signal Process.*, vol. 93, no. 6, pp. 1557–1565, 2013.

[10] J.-F. Connolly, E. Granger, and R. Sabourin, "An adaptive classification system for video-based face recognition," *Inf. Sci.*, vol. 192, pp. 50–70, Jun. 2012.

[11] M. Găianu and D. M. Onchiş, "Face and marker detection using Gabor frames on GPUs," *Signal Process.*, vol. 96, pp. 90–93, Mar. 2014.

[12] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosoph. Mag.*, vol. 2, no. 11, pp. 559–572, 1901.

[13] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[14] Y.-F. Liu, C.-Y. Lin, and J.-M. Guo, "Impact of the lips for biometrics," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3092–3101, Jun. 2012.

[15] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[16] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors," *Int. J. Comput. Vis.*, vol. 75, no. 2, pp. 247–266, 2007.

[17] *MIT CBCL Human Face Database*. [Online]. Available: http://cbcl.mit.edu/software-datasets/FaceData2.html, accessed Sep. 11, 2014.

[18] *INRIA Person Database*. [Online]. Available: http://pascal.inrialpes.fr/data/human, accessed Sep. 11, 2014.

[19] *NTUST MSPLab License Plate Database*. [Online]. Available: http://msp.ee.ntust.edu.tw/public%20file/LicensePlateDataset.rar, accessed Sep. 11, 2014.

[20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2008.

[21] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Proc. Pattern Recognit. Symp.*, vol. 2781. Sep. 2003, pp. 297–304.

[22] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *Advances in Biometrics* (Lecture Notes in Computer Science), vol. 4642. Berlin, Germany: Springer-Verlag, 2007, pp. 11–18.

[23] D. Abdullah, I. Murtza, and A. Khan, "Feature extraction and reduction strategy based on pyramid HOG and hierarchal exploitation of cortex-like mechanisms," in *Proc. 16th Int. Multi Topic Conf.*, Dec. 2013, pp. 160–165.
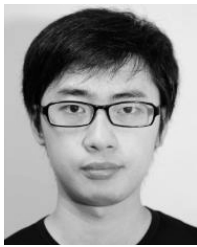
Yun-Fu Liu (S'09–M'13) received the master's degree in electrical engineering from Chang Gung University, Taoyuan, Taiwan, in 2009, and the Ph.D. degree in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2013.

He was involved in research with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA, USA, in 2012. In 2013, he joined the Multimedia Signal Processing Laboratory at the National Taiwan University of Science and Technology as a Post-Doctoral Fellow. He has worked on foreground segmentation, biometrics, digital halftoning, watermarking, image compression, and enhancement. His general interests lie in pattern recognition and image processing, and their related applications.

Dr. Liu was a recipient of the Doctoral Dissertation Excellence Awards from the Taiwanese Association for Consumer Electronics, the Institute of Information and Computing Machinery, and the Image Processing and Pattern Recognition Society of Taiwan, in 2013 and 2014, the Excellent Paper Award from the *Computer Vision, Graphics and Image Processing* in 2013, and the Master's Thesis Awards from the Taiwan Fuzzy Systems Association and ChiMei Optoelectronics in 2009.

Jing-Ming Guo (SM'10) received the Ph.D. degree from the Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, in 2004. He is currently a Professor with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei. His research interests include multimedia signal processing, biometrics, computer vision, and digital halftoning.

He is a fellow of the Institution of Engineering and Technology. He has been promoted to Distinguished Professor, in 2012, for his significant research contributions. He was a recipient of the Outstanding Youth Electrical Engineer Award from the Chinese Institute of Electrical Engineering in 2011, the Outstanding Young Investigator Award from the Institute of System Engineering in 2011, the Best Paper Award from the IEEE International Conference on System Science and Engineering in 2011, the Excellence Teaching Award in 2009, the Research Excellence Award in 2008, the Acer Dragon Thesis Award in 2005, the Outstanding Paper Awards from the Institute for Public Policy Research and *Computer Vision and Graphic Image Processing* in 2005 and 2006, and the Outstanding Faculty Award in 2002 and 2003.

Dr. Guo will be the General Chair of the IEEE International Conference on Consumer Electronics in Taiwan in 2015, and was the Technical Program Chair of the IEEE International Symposium on Intelligent Signal Processing and Communication Systems in 2012, the IEEE International Symposium on Consumer Electronics in 2013, and the IEEE International Conference on Consumer Electronics in Taiwan in 2014. He has served as a Best Paper Selection Committee Member of the IEEE TRANSACTIONS ON MULTIMEDIA. He was invited as a Lecturer of the IEEE Signal Processing Society's Summer School on Signal and Information Processing in 2012 and 2013. He was elected as the Chair of the IEEE Taipei Sections GOLD Group in 2012. He has served as a Guest Coeditor of two special issues of the *Journal of the Chinese Institute of Engineers* and the *Journal of Applied Science and Engineering*. He serves on the Editorial Board of the *Journal of Engineering*, *The Scientific World Journal*, *International Journal of Advanced Engineering Applications*, *Detection*, and the *Open Journal of Information Security and Applications*. He is currently an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, *Information Sciences*, and *Signal Processing*.

Chih-Hsien Hsia (M'10) was born in Taipei, Taiwan, in 1979. He received the M.S. degree in electrical engineering and the Ph.D. degree from Tamkang University, New Taipei, Taiwan, in 2005 and 2010, respectively.

He was a Visiting Scholar with Iowa State University, Ames, IA, USA, in 2007. From 2010 to 2013, he was a Post-Doctoral Research Fellow with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei. He joined the faculty of the Department of Electrical Engineering at Tamkeng University, from 2010 to 2013, as an Adjunct Associate Professor. He is currently an Associate Professor with the Department of Electrical Engineering, Chinese Culture University, Taipei. His research interests include DSP IC design, image/video compression system, multimedia signal processing, and computer/robot vision processing.

Dr. Hsia is a member of the Phi Tau Phi Scholastic Honor Society. He has served as a Guest Editor of the special issues for the *Journal of Applied Science and Engineering*. He serves on the Editorial Board of the *Journal of Control Science and Engineering*.

Sheng-Yao Su received the B.S. degree from the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan in 2013. He is currently pursuing the M.S. degree with the Institute of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include machine learning and computer vision.

Hua Lee (S'78–M'80–SM'83–F'92) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1974, and the M.S. and Ph.D. degrees from the University of California at Santa Barbara (UCSB), Santa Barbara, CA, USA, in 1978 and 1980, respectively. He returned to UCSB in 1990, where he is currently a Professor with the Department of Electrical and Computer Engineering. Before his return to UCSB, he was with the faculty of the University of Illinois at Urbana-Champaign, Champaign, IL, USA.

He served as the Department Vice Chair from 1993 to 1995 and the Department Chairman from 1998 to 2002. He currently serves as the Technical Director of the National Security Institute. His research interests cover the areas of imaging system optimization, high-performance image formation algorithms, synthetic aperture radar and sonar systems, acoustic microscopy, microwave nondestructive evaluation, terahertz imaging, tomographic ground-penetrating radar imaging, and dynamic sensing systems.

He served as the Chairman of the 18th, 24th, and 30th International Symposium on Acoustical Imaging in 1989, 1998, and 2009, respectively. He also served as the Cochair of the 13th International Workshop on Maximum Entropy and Bayesian Methods in 1993 and the 9th International Ground-Penetrating Radar Conference in 2002. From 1988 to 1994, he served as the Editor of the *International Journal of Imaging Systems and Technology*. He also served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1992 to 1995, and the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1994 to 1998.

Prof. Lee was a recipient of the Presidential Young Investigator Award for his work in imaging system optimization in 1985, the Professor of the Year Award from the Mortar Board National Honor Society in 1992, the 19th Pattern Recognition Society Award in 1993, the UCSB Academic Senate Distinguished Teaching Award in 1998, the UCSB College of Engineering Outstanding Faculty Award in 2000 and 2006, and the Technical Achievement Award of the International Symposium on Acoustical Imaging in 2011. He is a fellow of the Acoustical Society of America.