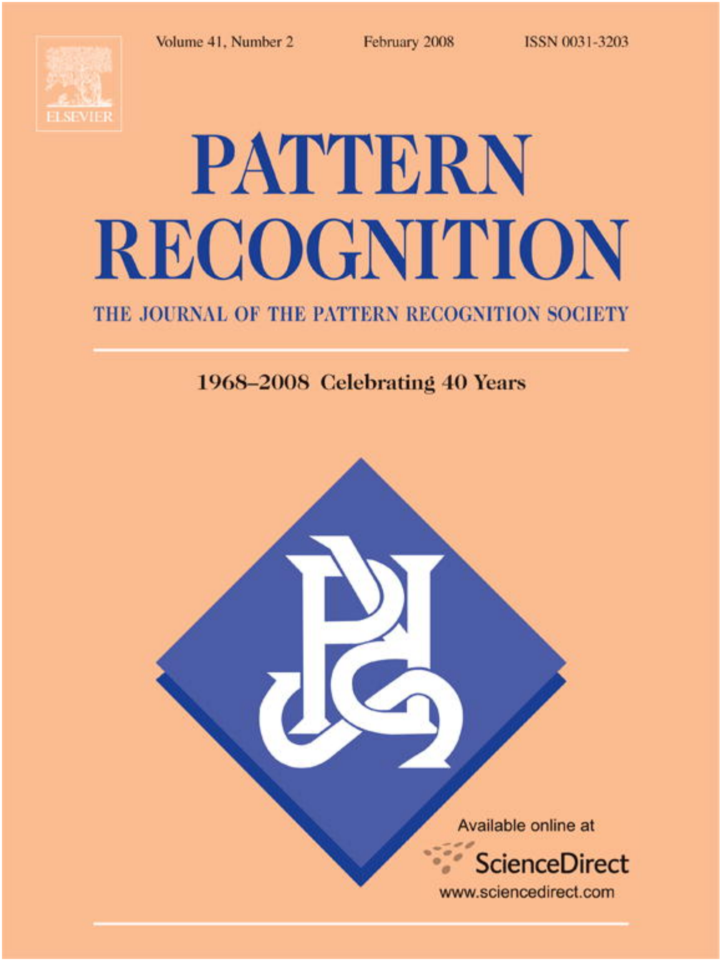
Volume 41, Number 2
February 2008
ISSN 0031-3203
ELSEVIER
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
1968–2008 Celebrating 40 Years
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 672–688
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Interest point detection using imbalance oriented selection

Qi Li[a,*], Jieping Ye[b], Chandra Kambhamettu[c]

[a]*Department of Computer Science, Western Kentucky University, USA*
[b]*Department of Computer Science and Engineering, Arizona State University, USA*
[c]*Department of Computer and Information Sciences, University of Delaware, USA*

Received 28 July 2006; received in revised form 3 May 2007; accepted 28 June 2007

Abstract

Interest point detection has a wide range of applications, such as image retrieval and object recognition. Given an image, many previous interest point detectors first assign interest strength to each image point using a certain filtering technique, and then apply non-maximum suppression scheme to select a set of interest point candidates. However, we observe that non-maximum suppression tends to over-suppress good candidates for a weakly textured image such as a face image. We propose a new candidate selection scheme that chooses image points whose zero-/first-order intensities can be clustered into two imbalanced classes (in size), as candidates. Our tests of repeatability across image rotations and lighting conditions show the advantage of imbalance oriented selection. We further present a new face recognition application—facial identity representability evaluation—to show the value of imbalance oriented selection.

Keywords: Interest point detection; Repeatability; Facial expression

1. Introduction

Interest point detection has a wide range of applications, such as image retrieval [1] and object recognition [2,3]. Many detectors have been presented before [2,4–15]. Most existing detectors contain the following two basic components: (i) assigning an interest strength to each image point, and (ii) selecting candidates.

In the previous work on interest point detection, great effort is focused on the study of strength (cornerness) assignments using different filtering techniques, such as first or second order Gaussian derivative [2], Gradient auto-correlation [4,10,16],and Laplace [17]. The authors of Ref. [18] introduced unit step edge function (USEF) to model straight line edges, and proposed an accurate and flexible multi-corner detector. Interest point detection can also be formulated as an optimization problem, and then solved by Genetic Programming [19,20].

Non-maximum suppression is a popular method for selecting good candidates in existing detectors [2,4,10]. Assume each image point has been assigned an interest strength. Non-maximum suppression resets the strength of a point to zero, i.e., eliminates its candidacy, if it is not a local maximum. With non-maximum suppression, interest points tend to scatter in the entire image plane. This may be desirable for highly textured images that contain rich textures in the entire image plane, but might not be desirable for weakly textured images such as face images. In the latter situation, non-maximum suppression over-suppresses good candidates. Another issue of non-maximum suppression is that it can destroy local geometry information, as illustrated in Section 2.1. The authors of Ref. [21] proposed *adaptive* non-maximum suppression to obtain spatially well distributed interest points over images. Instead of using a fixed-size suppression window, adaptive non-maximum suppression dynamically decreases the size of the suppression window. Adaptive non-maximum suppression has been shown to be competitive to the standard one in image mosaicing. To address the nature of weak textures in face images, the authors of Refs. [22,23] used the strategy of thresholding strength maps without any candidate selection scheme, where strength maps are computed using entropy and local variance, respectively. The number of local image patches extracted from a face image is thus very large, which has two

* Corresponding author. Tel.: +1 302 266 0586.
E-mail addresses: qi.li@wku.edu (Q. Li), jieping.ye@asu.edu (J. Ye), chandra@cis.udel.edu (C. Kambhamettu).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.06.020

limitations: (i) expensive cost in matching image patches; and (ii) higher probability of mismatching.

In this paper, we propose an imbalance oriented scheme that chooses image points whose zero-/first-order intensities are clustered into two imbalanced classes (in size), as candidates. Our basic motivation for imbalance oriented selection is to minimize the occurrences of edge points. It is worth noting that edge points are usually not considered as good features since they have similar local appearances, which increases uncertainty in matching local appearances. Besides the increasing ambiguity, the number of edge points in an image is usually large, which can result in high computational cost in high-level applications such as recognition. Many existing interest point detectors involve certain mechanisms to penalize edge response. For example, the Harris detector involves a penalty term in the strength assignment to penalize the edge response [4,10]; DoG applies second-order intensity information, determined by a Hessian matrix, to verify edge response and penalize edge points [2]. Our rationale is that edge points can be characterized as points of balanced local appearances, and they can be minimized via an imbalance criterion.

Without involving any suppression window, imbalance oriented selection can output interest points "non-uniformly", i.e., it may detect a larger number of points in certain regions than other regions, and thus provide a wise alternative for non-maximum suppression. In the literature, SUSAN [8] provides an implementation for zero-order imbalance selection. Based on a simple binary clustering approach, we propose a unified scheme for imbalance oriented selection using zero or first order local information. We combine three popular interest assignments (Gradient auto-correlation, Gradient, and Laplace) with three candidate selection schemes (non-maximum suppression, zero- and first-order imbalance oriented selection), and derive nine detectors. We perform repeatability tests, first proposed by Schmid et al. [10], on these detectors, under two different image variations: rotations and lighting conditions. Our results show the advantage of imbalance oriented selection. For example, in the test of repeatability across lighting conditions, the average repeatability rate achieved by imbalance oriented selection is 20% higher than the one achieved by non-maximum suppression.

We also present a new face recognition application to show the value of imbalance oriented selection. Note that face images are weakly textured. Most previous studies in face recognition focused on how to represent appearance instances. Little attention, however, was given to the problem of how to select "good" instances for a gallery, which may be called the *facial identity representation problem*. We will give an evaluation of the identity representability of facial expressions. The *identity representability* of an expression is measured by the recognition accuracy achieved by using its samples exclusively as the gallery data. Feature distributions are used to represent appearance instances. A feature distribution of a face image is based on the number of occurrences of interest points in regular grids of an image plane. Our study shows that certain facial expressions, such as neutral ones, have stronger identity representability than other expressions. (For convenience, we consider neutral as a facial expression.)

The rest of the paper is organized as follows: Section 2 first describes non-maximum suppression, and then gives a unified scheme for imbalance oriented selection. Section 3 presents nine detectors. Section 4 presents repeatability tests. Section 5 presents the evaluation of identity representability. Finally, conclusions are given in Section 6.

2. Candidate selection schemes

Candidate selection schemes aim to select a set of reliable candidates for the final interest point output. To decide the candidacy of an image point, a selection scheme may or may not use its interest strength. A popular scheme is maximum oriented selection, such as non-maximum suppression, which will be described in Section 2.1. Unlike maximum oriented selection, imbalance oriented selection is based on zero- or first-order imbalanced local information of a point, as described in the Sections 2.2 and 2.3, respectively.

2.1. Maximum oriented selection

Maximum oriented selection has been widely used in previous detectors [2,4,8,10]. A typical implementation of maximum oriented selection is non-maximum suppression. To decide the candidacy of a point p, non-maximum suppression checks the following condition:

$$\text{strengh}(p) > \text{strength}(q) \quad \forall q \in O_p \backslash \{p\},$$

where O_p is a small neighborhood of p (usually a 3×3 window centered by p [10]). If the condition is false, then strength(p) is reset to zero.

A limitation of non-maximum suppression is that it may destroy the geometric structure of a local appearance, as shown in Fig. 1, where a number indicates the strength of the associated image point. The solid black circles have large strength and are good candidates for interest points; the circles are not. When non-maximum suppression is applied (using a 3×3 suppression window, indicated as a dashed square), the only possible candidate in each case is the image point with strength 10, which fails to reveal the difference in geometric structures in the two cases.

Furthermore, non-maximum suppression tends to over-suppress many valuable candidates if an image is weakly textured. Adaptive non-maximum suppression was recently proposed to address these issues in non-maximum suppression, where the basic goal is to achieve better distributed interest points. More details can be found in Ref. [21].

2.2. Zero-order imbalance

As the name implies, zero-order imbalance selection decides the candidacy of an image point p by measuring the imbalance of the zero-order local intensity information. SUSAN [8] is an implementation of this scheme. Given a point p and its

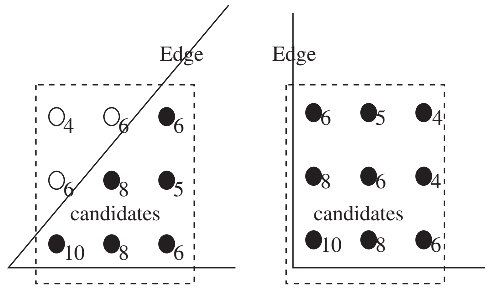

Fig. 1. Non-maximum suppression destroys the geometric structure.

circular neighborhood O_p, SUSAN first separates the O_p into two clusters, by measuring the condition $|I(q)-I(p)| \leqslant t_1, q \in O_p$, where t_1 is a threshold. Then it basically computes the ratio of the sizes of these two clusters (small size over large size), denoted as r_p. The point p with $r_p \leqslant t_2$ is chosen as a candidate, where t_2 is the imbalance threshold.

We propose an implementation of zero-order imbalance selection by first sorting $\{I(q)|q \in O_p\}$ (in increasing order), and then looking for the intensity whose difference with the next intensity is maximum. The rank of this intensity will be called the *index of the maximum difference*. The index of the maximum difference will be used to split the set of intensities into two clusters later. The implementation includes a step of discarding noisy points. There are two motivations for this formulation. First, defining zero-order imbalance of a patch may not be necessarily dominated by the intensity of the center point; second, the clustering method can be naturally extended to define first-order imbalance, as described in Section 2.3. The implementation is described in *Algorithm 1*.

In all experiments in this paper, we choose threshold1 $=10$, and threshold2 $=0.45$. A 3×3 window is used to define the neighborhood points. We found that zero-order balance is more appropriate for interest point detection in weakly textured images than highly textured ones.

Algorithm 1. Zero-order imbalance (proposed version)
1. For each image point p
2. Sort the intensities of neighboring points of p
3. Find the maxDiff and the index of maximum difference
4. If maxDiff < threshold1
5. strength(p) $\leftarrow 0$ // discarding noisy points
6. Else
7. Compute r_p
8. If $r_p >$ threshold2
9. strength(p) $\leftarrow 0$ // discarding edge points

2.3. *First-order imbalance*

This scheme uses the first-order intensity information of an image point p to decide its candidacy. In Ref. [15], we proposed an implementation of first-order imbalance that first clusters the magnitudes of first-order derivatives of p along n directions into two classes: strong and weak, and then selects p as a candidate only if p has less than $n/2$ strong first-order changes. The clustering method used in Ref. [15] is based on the mean

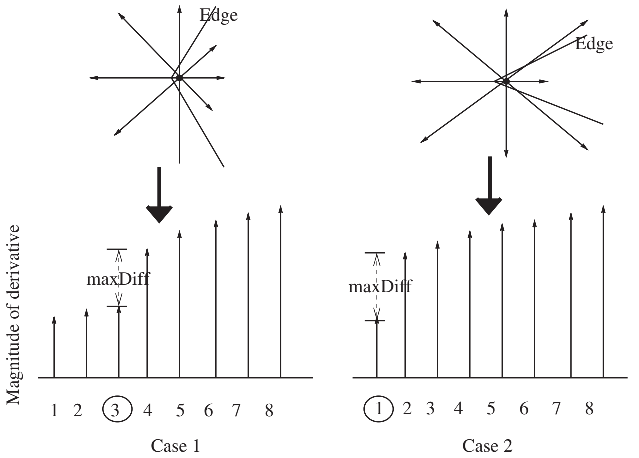

Fig. 2. Two cases indicating desirable interest points and the sorted intensity changes. The indices of maximum difference in the two cases are 3 and 1, respectively.

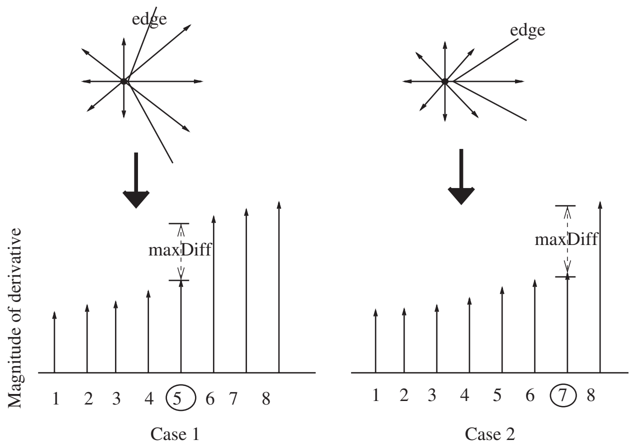

Fig. 3. Discarding redundant interest points. The indices of maximum difference in the two cases are 5 and 7, respectively.

change, i.e., the change that is larger (smaller) than the mean change is considered to be strong (weak).

Using the same clustering method in Algorithm 1, we propose first-order imbalance oriented selection implementation, as described in *Algorithm 2*. Fig. 2 gives an illustration of two cases of selecting desirable candidates, where we consider $n=8$ directions. Intensities are supposed to change slightly at the same side of an edge. A long (short) arrow indicates a strong (weak) intensity change along the associated direction. Fig. 3 gives illustrations of discarding redundant points that are the co-occurrence of certain desirable points shown in Fig. 2. Fig. 4 gives illustrations of edge points and noisy points.

The threshold in discarding noisy points is the only significant parameter for the first-order imbalance oriented selection. In all experiments in this paper, it is set to 0.5. We have found that the parameter setting is not sensitive to input images.

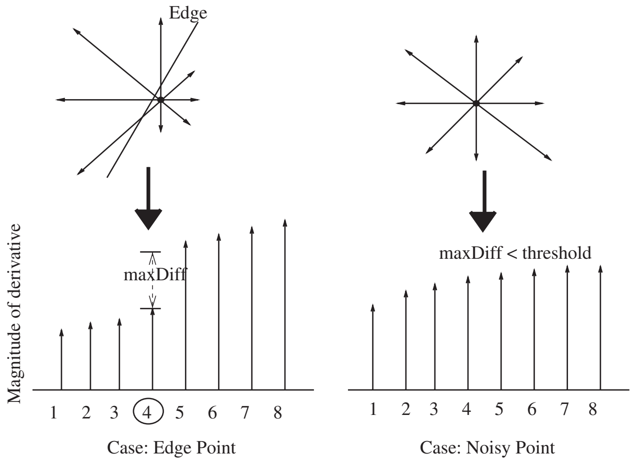

Fig. 4. Discarding an edge point and a noisy point. The index of maximum difference of the edge point is 4.

Algorithm 2. First-order imbalance

1. For each image point p
2. For each of n directions
3. Compute the change along that direction (by convolving with the first derivative of a Gaussian)
4. Sort the n changes
5. Find the maxDiff and the index of maximum difference
6. If maxDiff $<$ threshold
7. strength(p) $\leftarrow 0$ // discarding noisy points
8. Else
9. If the index $\geqslant n/2$
10. strength(p) $\leftarrow 0$ // discarding edge points and "redundant" interest points

3. Constructing interest point detectors

A detector can be formulated as a combination of a strength assignment with a selection scheme, in addition to a thresholding step. For example, the Harris detector is derived by combining the gradient auto-correlation assignment and non-maximum suppression. We will construct nine detectors by combining the above three selection schemes with the following three strength assignments:

- *Gradient auto-correlation*: The interest strength of an image point p is computed by the eigenstructure of p's gradient auto-correlation matrix

$$C(p) = \begin{pmatrix} \sum_{q \in O_p} w_q I_x^2(q) & \sum_{q \in O_p} w_q I_x(q) I_y(q) \\ \sum_{q \in O_p} w_q I_x(q) I_y(q) & \sum_{q \in O_p} w_q I_y^2(q) \end{pmatrix},$$

where I_x and I_y are the horizontal and vertical derivatives, respectively, O_p is a neighborhood of p, and $(w_q)_{q \in O_p}$ is a smoothing filter. In the Harris detector [4], the interest strength of a point p is defined to be the summation of the eigenvalues of $C(p)$. To reduce the computational cost, the following equivalent form is more commonly used in practice [10]:

$$\text{strength}(p) = \det(C(p)) - \alpha \text{trace}^2(C(p)), \tag{1}$$

where α is a discriminant factor that is usually set to 0.06 [10].

- *Gradient*: Given a point p, the magnitude of its gradient $[I_x(p), I_y(p)]$ can be assigned as the interest strength of p. In this paper, we use an alternative, i.e., the largest of the magnitudes of first-order derivatives to p as the interest strength, i.e.,

$$\text{strength}(p) = \text{largest change}(p). \tag{2}$$

- *Laplace*: Given an image point p, the Laplace assigns the strength to p as follows [17]:

$$\text{strength}(p) = |I_{xx}(p) + I_{yy}(p)|. \tag{3}$$

A close approximation of Laplace is DoG (Difference of Gaussian) [2]. More details on Laplace and DoG can be found in Refs. [2,17].

3.1. *Combining a strength assignment and a candidate selection scheme*

Fig. 5 shows the interest strength maps of a face instance, associated with three assignments, before and after three different selections are applied. Let us first consider the behavior of three assignments by comparing the three maps in the first column (i.e., before a selection). Gradient auto-correlation appears to localize potential interest points at the region level, illustrated by the isolated white regions in the top figure. In contrast to gradient auto-correlation, gradient and Laplace give attention to edges.

Now let us consider the behavior of three selections. We can observe that non-maximum suppression tends to sample candidates in the entire image plane. Zero-order imbalance and first-order imbalance appear to thin the isolated regions or edges in the original strength maps. But clearly, the thinning effectiveness of first-order imbalance is much more considerable than the effectiveness of zero-order imbalance. Typically, first-order imbalance significantly reduces the number of potential interest points along the head boundary, which is important for a detector to extract local descriptors of a face image taken with a different background.

For a weakly textured image, it may be fair to say that non-maximum suppression tends to over-suppress the candidates of interest points, and zero-order imbalance tends to over-select the candidates. We will observe that first-order imbalance outperforms these two selection schemes, in the repeatability evaluation.

3.2. *Thresholding*

As shown in Fig. 5, the number of candidates may be large, even for non-maximum suppression selection. Many of them may contain similar local appearance (after all, a face image is weakly textured), which increases the chance of mismatching

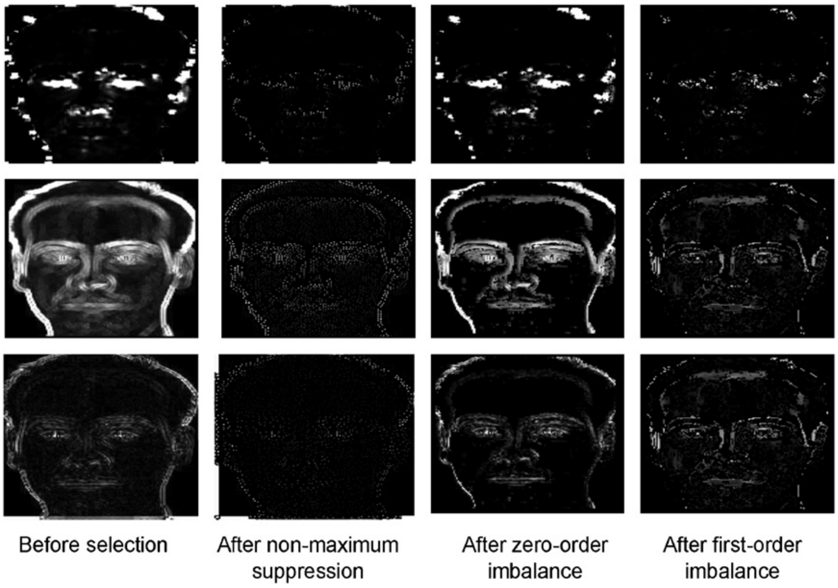

Fig. 5. Interest strength maps under three assignment schemes, before and after candidate selection. First row, gradient auto-correlation; second row, gradient; third row, Laplace. The values are scaled for visualization enhancement.

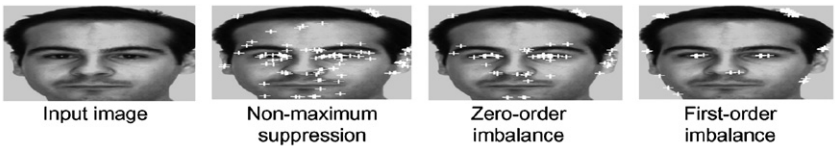

Fig. 6. One hundred interest points in a face image. Gradient auto-correlation is used as strength assignment.

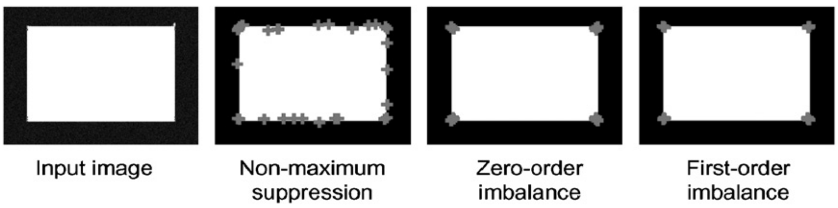

Fig. 7. Thirty interest points in the image of a white board with a black background. Gradient auto-correlation is used as strength assignment.

in applications. Furthermore, a large number of interest point outputs can significantly increase the computational cost in applications.

Two *thresholding* strategies can be used to determine the final output of interest points from a candidate set, after being sorted according to the strength. One strategy is to threshold the strength, and the other is to threshold the number of outputs. In our study, we choose the second one as thresholding the number of outputs is convenient under different strength assignment schemes.

Fig. 6 shows on the output of interest points via different candidate selection schemes, using gradient auto-correlation as the strength assignment. We can observe that non-maximum suppression leads to the output of interest points scattered in a large number of image regions while imbalance oriented selection schemes leads to the output of interest points gathered in a few numbers of image regions. This observation can be further justified by a comparison of the capability of the selection schemes in suppressing edge points. Let us consider the following simple example, a white board with a black background, as shown in Fig. 7. In this circumstance, the edge points on the four boundaries of the white board are expected to have distinctly larger strength than other points. Furthermore, due to the reality of the existence of noise, the strength of an edge point may be slightly different from the strength of its neighboring edge points. (The image of a white board in Fig. 7 contains 0.1% noise.) So, after non-maximum suppression, the output of interest points will consist of a subset of edge points scattered around the four boundaries. However, with an imbalance oriented scheme, the output of interest points will only consist of a few image points around the four corners of the white board, due to the stronger ability of the scheme in suppressing edge points.

4. Repeatability evaluation

In this section, we test the repeatability of interest point detectors under two kinds of transformations: rotations and lighting conditions. More specifically, our focus is on the comparison between non-maximum suppression and imbalance oriented selection schemes. We follow the definition of ε-repeatability rate proposed in Ref. [10], where ε, in pixel units,

Table 1
Three parameters for repeatability across rotations

Parameter	Value
Rotation angles	$\theta_i = -45° + (i-1) \times 10,\ i = 1 \ldots 10$
Tolerance	$\varepsilon_i = 0.5 + (i-1),\ i = 1 \ldots 5$
Number of interest points	$n_i = 140 + (i-1) \times 20,\ i = 1 \ldots 14$

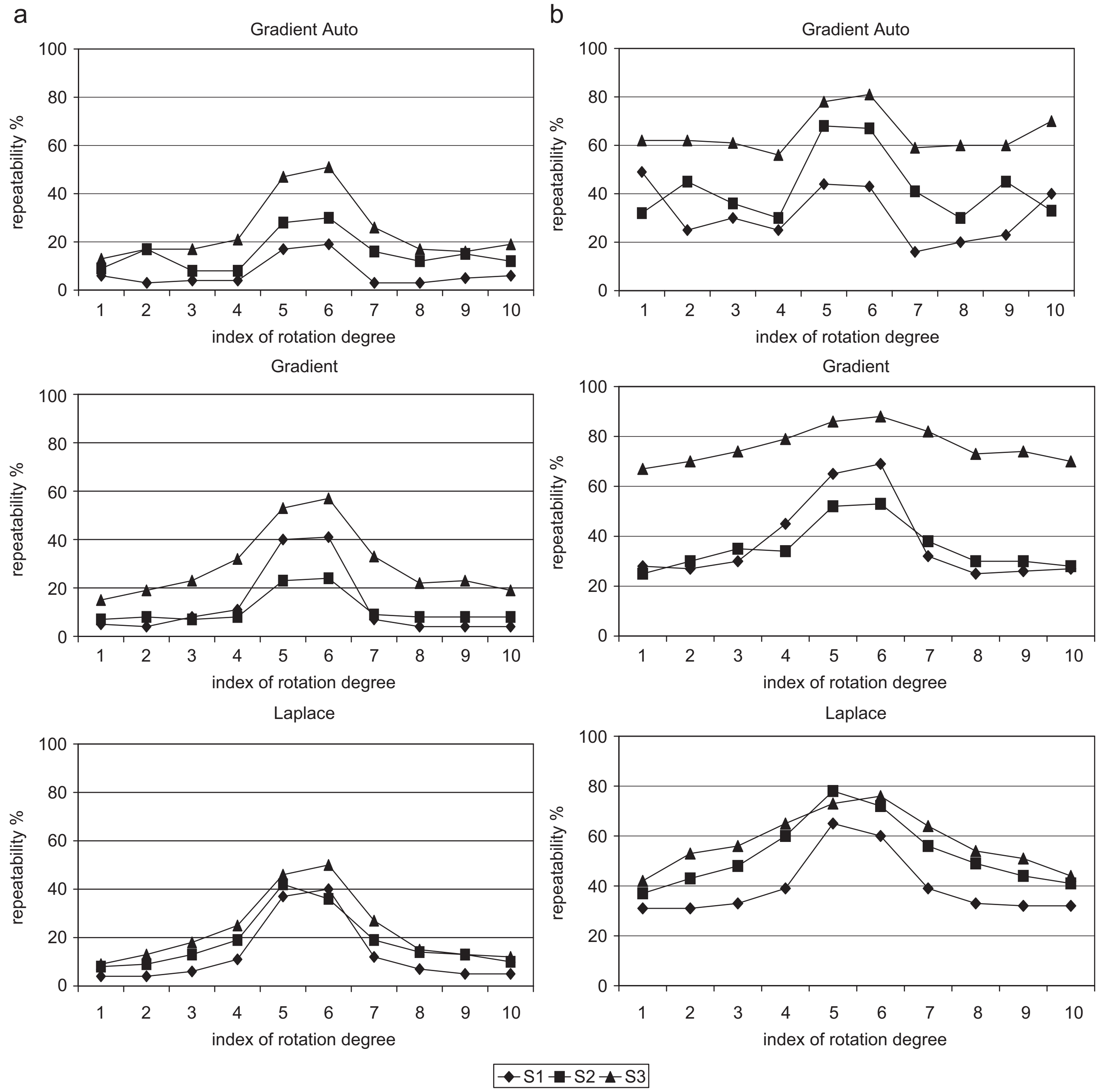

Fig. 8. ε-repeatability rates of nine detectors across rotations (from $\theta_1 = -45°$ to $\theta_{10} = 45°$). S1, non-maximum suppression; S2, zero-order imbalance; and S3, first-order imbalance. Note that GradientAuto/S1, Harris. (a) $\varepsilon_1 = 0.5$ and (b) $\varepsilon_2 = 1.5$.

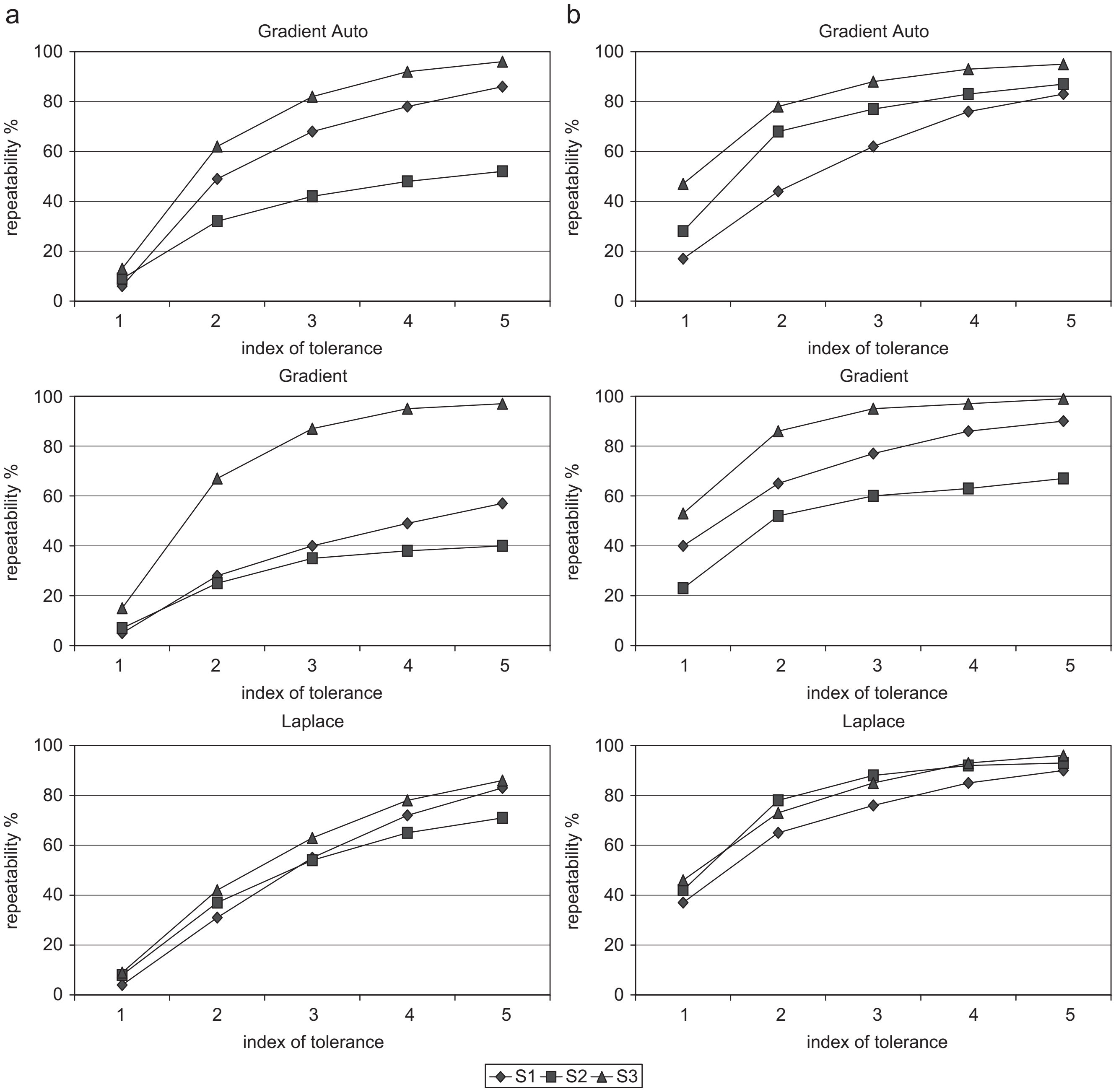

Fig. 9. ε-repeatability rates of nine detectors with respect to localization tolerances (from $\varepsilon_1 = 0.5$ to $\varepsilon_5 = 4.5$): (a) $\theta_1 = -45°$ and (b) $\theta_5 = -5°$.

is the tolerance in locating interest points under a transformation. For example, the transformation used in evaluation of repeatability across lighting condition is an identity transformation since an interest point detected under one lighting condition is expected to be detected in the same location under other lighting conditions.

In the zero-order imbalance algorithm, the ratio threshold is set to 0.5, and a 3×3 window is used to define the neighboring points. In the first-order imbalance algorithm, we choose $n = 8$ in the implementation (n is the number of directions), as shown in the illustration in Fig. 2. (We observed that the involvement of many interpolation operations with $n > 8$ brings substantially higher computation cost, while not improving the repeatability rate.)

For convenience, we introduce the conventions: S1 = non-maximum suppression, S2 = zero-order imbalance, and S3 = first-order imbalance. Furthermore, we use the following convention in naming nine detectors: Strength Assignment/Si. For example, GradientAuto/Si indicates the detector constructed by gradient auto-correlation and Si selection scheme.

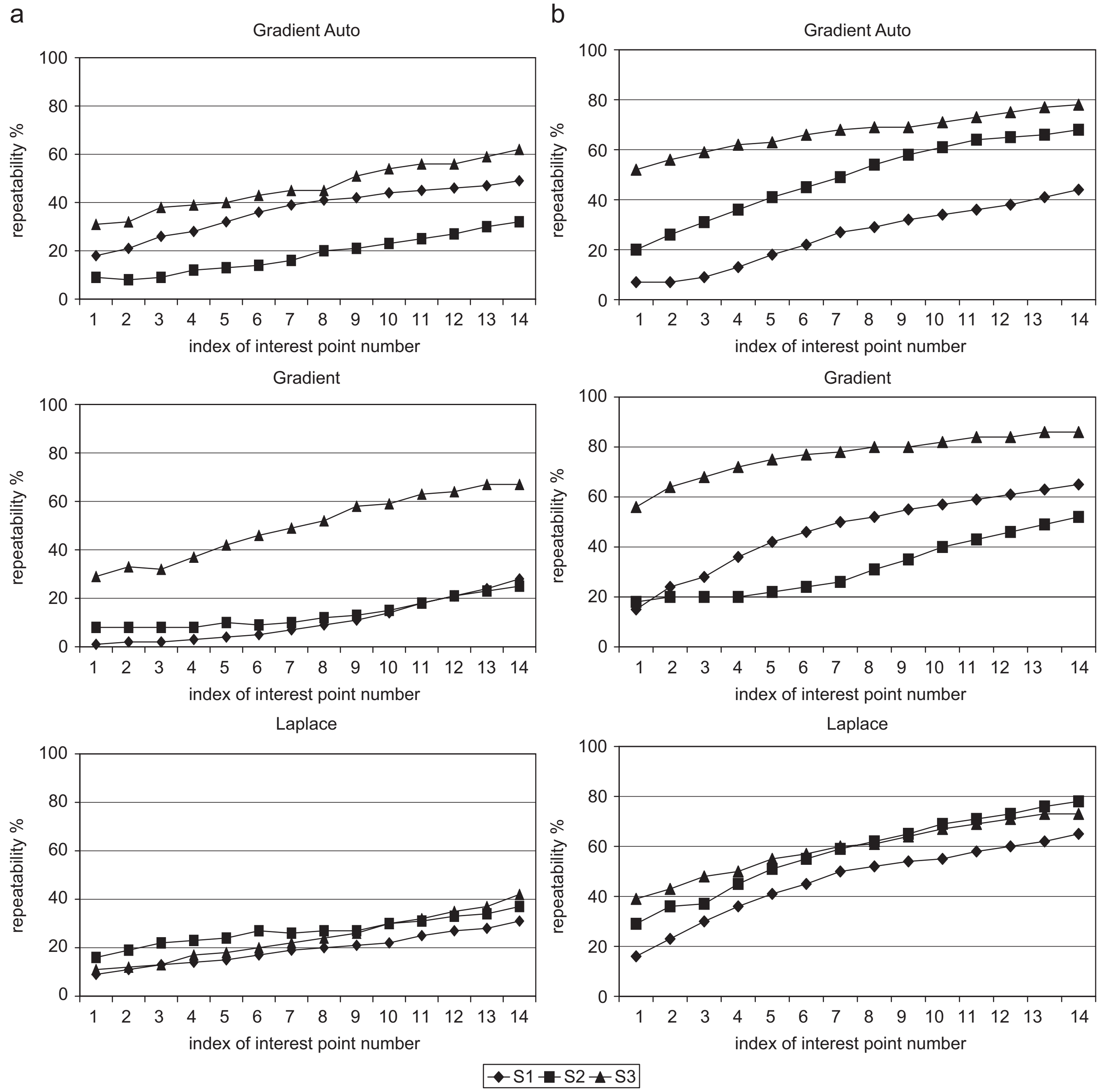

Fig. 10. $\varepsilon_2 = 1.5$-repeatability rates of nine detectors with respect to interest point numbers (from $n_1 = 140$ to $n_{14} = 400$): (a) $\theta_1 = -45°$ and (b) $\theta_5 = -5°$.

4.1. Repeatability rate across rotations

Similar to the design in Ref. [2], a rotated image is generated synthetically. Nearest-neighbor backward rendering is used to generate a rotated image. The dimension of the original face image is 142×180. We will present the results on repeatability across rotation angles, in addition to the results on repeatability with respect to localization tolerance and the number of output interest points. Table 1 lists the notations of three relevant parameters.

Repeatability across rotation θ_i: We limit the output of interest points to $n_{14} = 400$. Fig. 8 shows the repeatability rates across θ_i, from $\theta_0 = -45°$ to $\theta_{10} = 45°$. Fig. 8(a) and (b) are associated with $\varepsilon_1 = 0.5$ and $\varepsilon_2 = 1.5$ pixels, respectively. Three assignments appear to be competing with each other. Among three selections, first-order imbalance usually outperforms the others. The behavior of zero-order imbalance is interesting; it appears to have "affinity" towards gradient auto-correlation and Laplace assignments that make it outperform non-maximum suppression. Among the nine detectors, Gradient/S3 performs better than the others.

Repeatability with respect to ε_i: We consider two rotation parameter settings: a large rotation $\theta_0 = -45°$, and a small rotation $\theta_5 = -5°$. The output of interest points is limited to

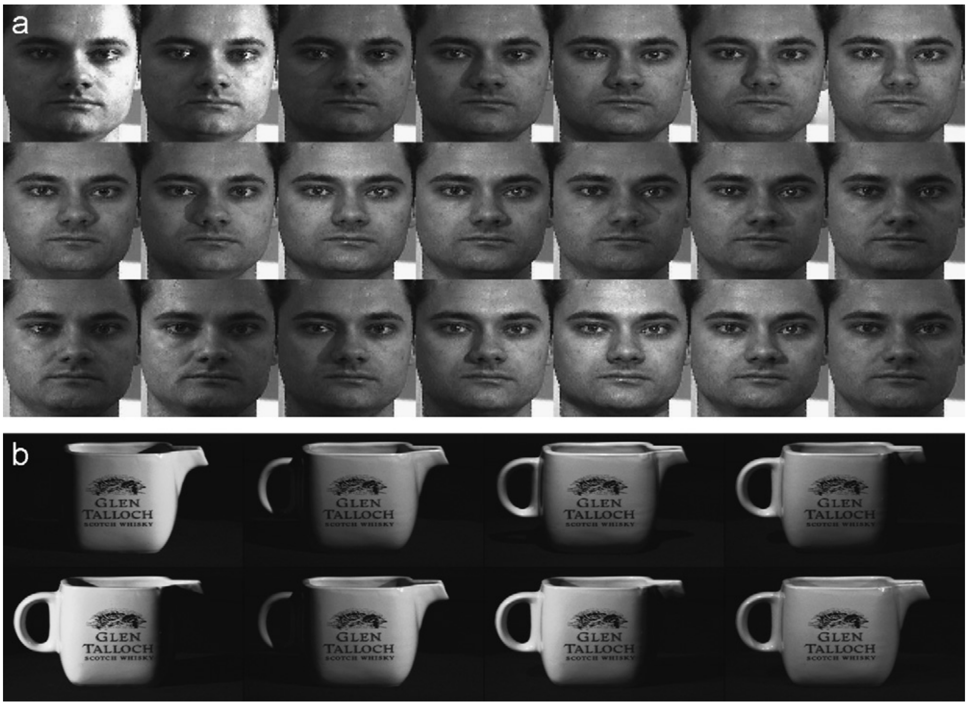

Fig. 11. Two data sets for repeatability evaluation across lighting conditions: (a) CMU-PIE 21 lighting conditions and (b) ALOI 8 lighting conditions.

$n_{14} = 400$. Fig. 9 shows the repeatability rates with respect to ε_i, from $\varepsilon_1 = 0.5$ to $\varepsilon_5 = 4.5$. Fig. 9(a) and (b) are associated with rotations $-45°$ and $-5°$, respectively. In most cases, first-order imbalance shows better performance than others. Non-maximum suppression and zero-order imbalance are competing with each other—their performance depends not only on the assignments but also on the rotation angles.

Repeatability with respect to n_i: We consider two rotation parameter settings: a large rotation $\theta_0 = -45°$, and a small rotation $\theta_5 = -5°$. The tolerance parameter is set to $\varepsilon_2 = 1.5$. Fig. 10 shows the repeatability rates, with respect to n_i, $i = 1 \ldots 14$. From Fig. 10, we can observe that, for the majority of detectors, increasing the output n_i is effective in lifting up the repeatability rates. In most cases, first-order imbalance outperforms the other two selection schemes. Zero-order imbalance, combined with gradient auto-correlation or Laplace, is somehow competing with the first-order imbalance if the rotation angles are small. Among the nine detectors, gradient/S3 performs the best.

From the tests on other face images, we observed similar results. As a summary, our main observation is that, for face images, the first-order imbalance achieves better performance than the other two selections in this evaluation.

4.2. Repeatability across lighting conditions

We now evaluate the repeatability of the nine detectors across lighting conditions on two different data sets: (i) CMU-PIE 21 lighting conditions [24]; and (ii) ALOI 8 lighting conditions [25], as shown in Fig. 11(a) and (b), respectively. The dimensions of CMU-PIE images are 220×175, and these images are manually aligned using the raw images in CMU-PIE. The dimensions of ALOI images are 288×384.

For convenience, these images (PIE and ALOI) are indexed from left to right and top to bottom. It is worth noting that the 0th face instance (i.e., the top leftmost one) is illuminated by an extreme light source, which makes its appearance significantly different from other instances. In the later repeatability test, we will always compute the repeatability rate between 0th image and the others.

We use similar parameter setting and presentation of results as given in Section 4.1. We present repeatability rates across lighting conditions, with respect to localization tolerance, and with respect to interest point numbers.

On CMU-PIE images, Fig. 12 shows the repeatability rates associated with indices i, from 1 to 20; Fig. 13(a) and (b) show the repeatability rates with respect to tolerance and number of interest points, respectively, where we consider the pair of lighting conditions (0th, 10th). In Fig. 13(b), ε is set to 1.5.

On ALOI images, Fig. 14 shows the repeatability rates associated with indices i, from 1 to 7; Fig. 15(a) and (b) show the repeatability rates with respect to tolerance and number of interest points, respectively, where we consider the pair of lighting conditions (0th, 7th). In Fig. 15(b), ε is set to 1.5.

On both data sets, it is clear that both imbalance oriented selection schemes usually outperform the non-maximum suppression. For example, with tolerance $\varepsilon = 1.5$, the average repeatability rates achieved by imbalance schemes on CMU-PIE (ALOI) are about 20% (40%) higher than by non-maximum suppression. It is interesting to note that two imbalance selection schemes on the lighting test are competitive to each other. There is an exception on performance of the combina-

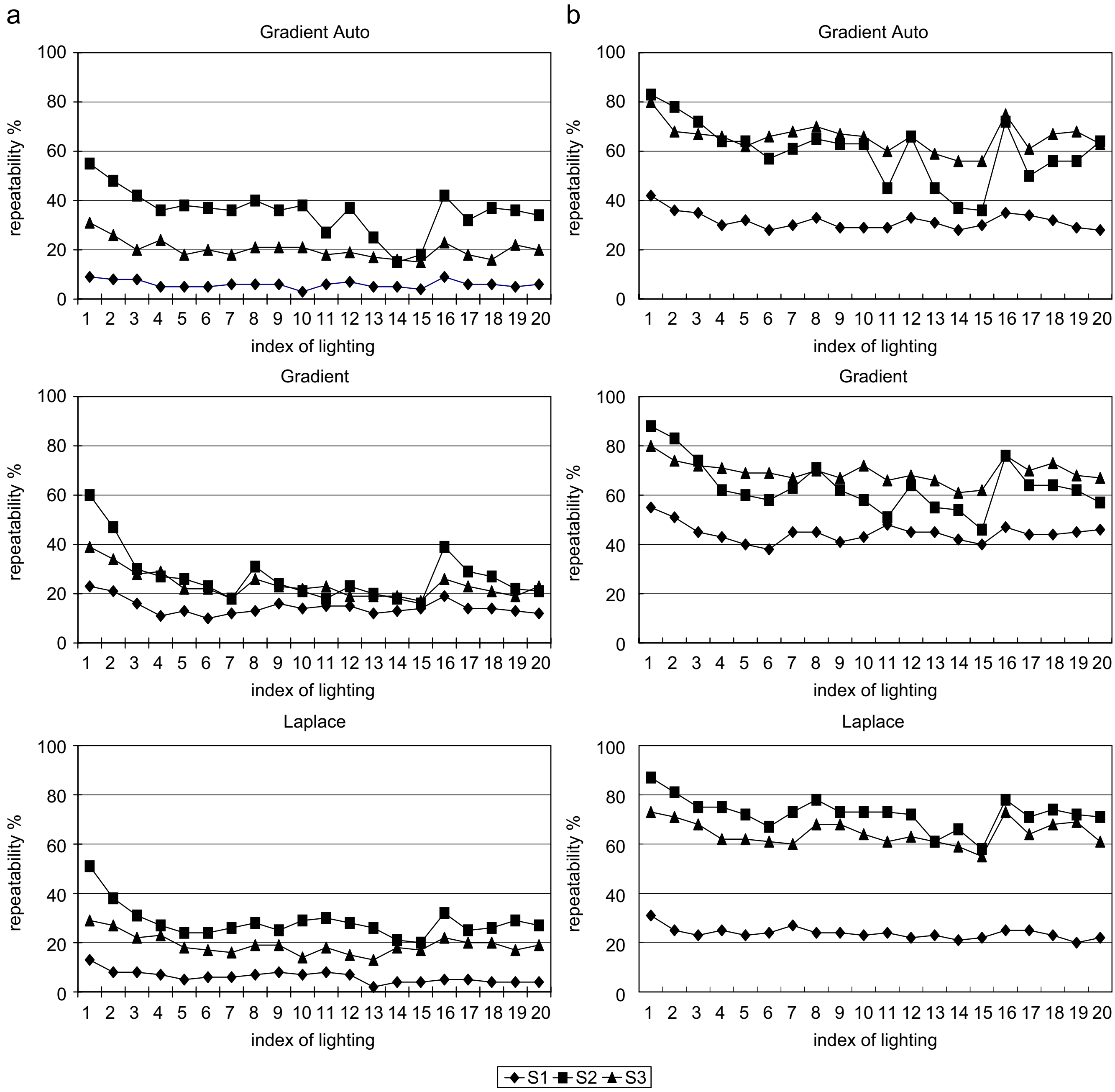

Fig. 12. CMU-PIE: ε-repeatability rates of nine detectors across lighting conditions: (a) $\varepsilon_1 = 0.5$ and (b) $\varepsilon_2 = 1.5$.

tion of gradient strength assignment and zero-order imbalance oriented selection, that is, gradient degrades the performance of zero-order imbalance (as shown in Fig. 14). This may imply the difference in the nature of the appearance of a face and a cup. This observation can somehow explain the fact that gradient auto-correlation and Laplace are much more popular than gradient in the literature on interest point detection.

We tested other pairs of lighting conditions, and furthermore images of other objects in the two data sets. We obtain similar results on the repeatability evaluation. We also tested DoG method used in Ref. [2] on our data, which achieved comparable repeatability to Harris, one of the nine tested detectors. Thus, based on our comprehensive tests, we conclude the superiority of imbalance oriented selection schemes over non-maximum suppression. From the evaluation results, we also observe that the first-order imbalance scheme achieves a higher repeatability rate than zero-order imbalance in most cases. Recall that first-order imbalance involves only one parameter that is used to suppress noisy points. We set that parameter to be 0.5 in this paper. But we had tried a range of parameter value from 0.5 to 2, and found that the variation of the repeatability rate is usually less than 5%. Zero-order imbalance involves two parameters: threshold1 is used to suppress noisy points, and threshold2 is used to suppress edge points. It was found that these parame-

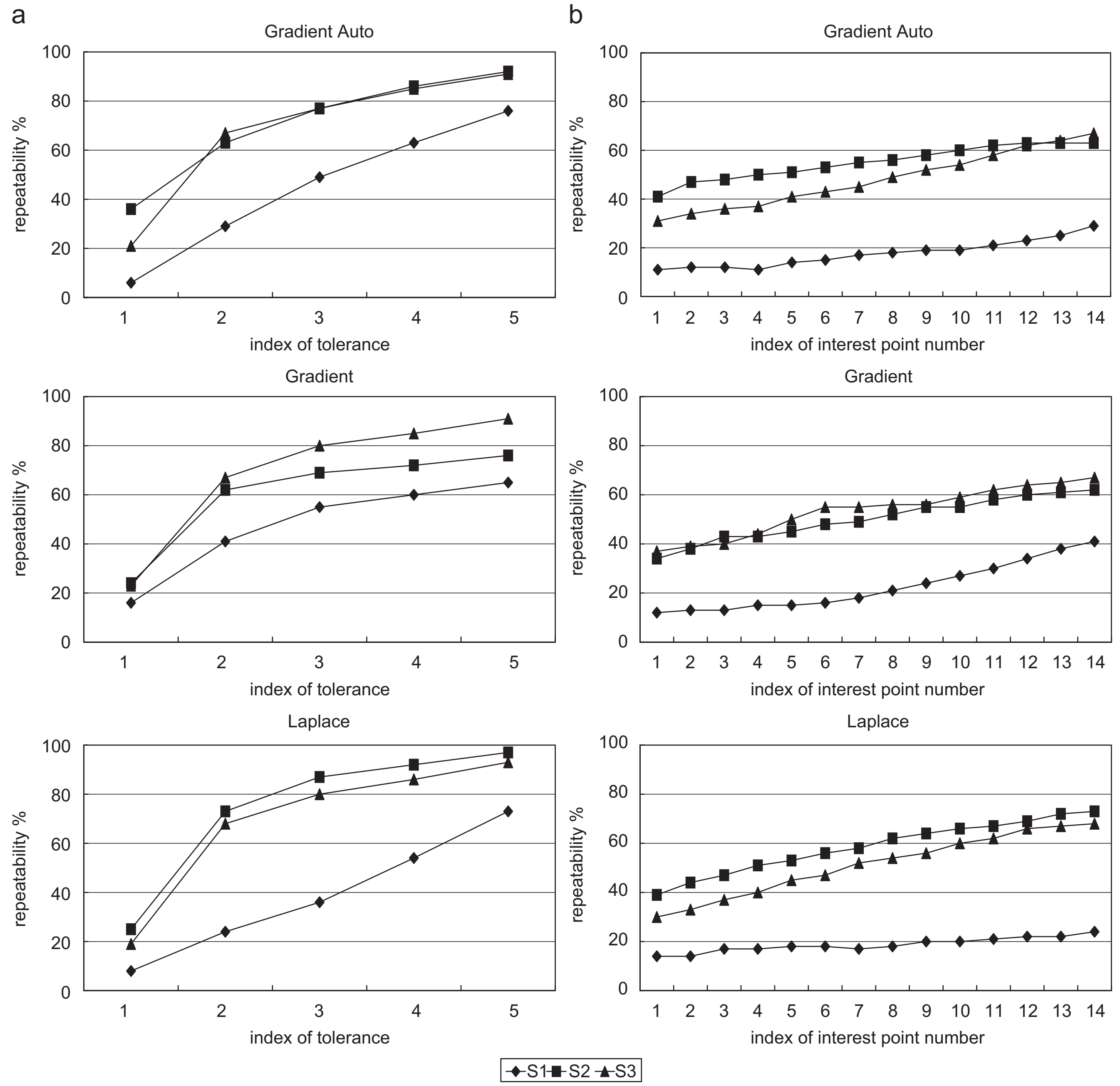

Fig. 13. CMU-PIE: (a) ε-repeatability rates of nine detectors with respect to localization tolerances (from $\varepsilon_1 = 0.5$ to $\varepsilon_5 = 4.5$); and (b) $\varepsilon_2 = 1.5$-repeatability rates of nine detectors with respect to interest points numbers (from $n_1 = 140$ to $n_{14} = 400$).

ters are relatively more sensitive to input images than the parameter in first-order imbalance. The variation of repeatability rate may be over 5% with a slight change on a parameter (e.g., threshold2 is changed from 0.45 to 0.46). These facts show the better robustness of first-order imbalance.

4.3. Discussion on richly textured images

We have addressed the issue of non-maximum suppression in detecting interest points in weakly textured images. In the following, we give a discussion on richly textured images, which can give us a more comprehensive view on non-maximum suppression and imbalance oriented selection.

Fig. 16 shows the image Van Gogh's sower that is used in the evaluation of interest point detectors by Schmid et al. [10]. The sower image is richly textured. We apply Harris detector to the sower image under different levels of noise, ranging from 2% to 10% (of 255), and obtain the repeatability with respect to noise. Similarly, we apply Harris detector to the face image used in Section 4.1 under different levels of noise. The results are shown in Fig. 17. It is clear that the repeatability on the sower image is much more robust with respect to noise than

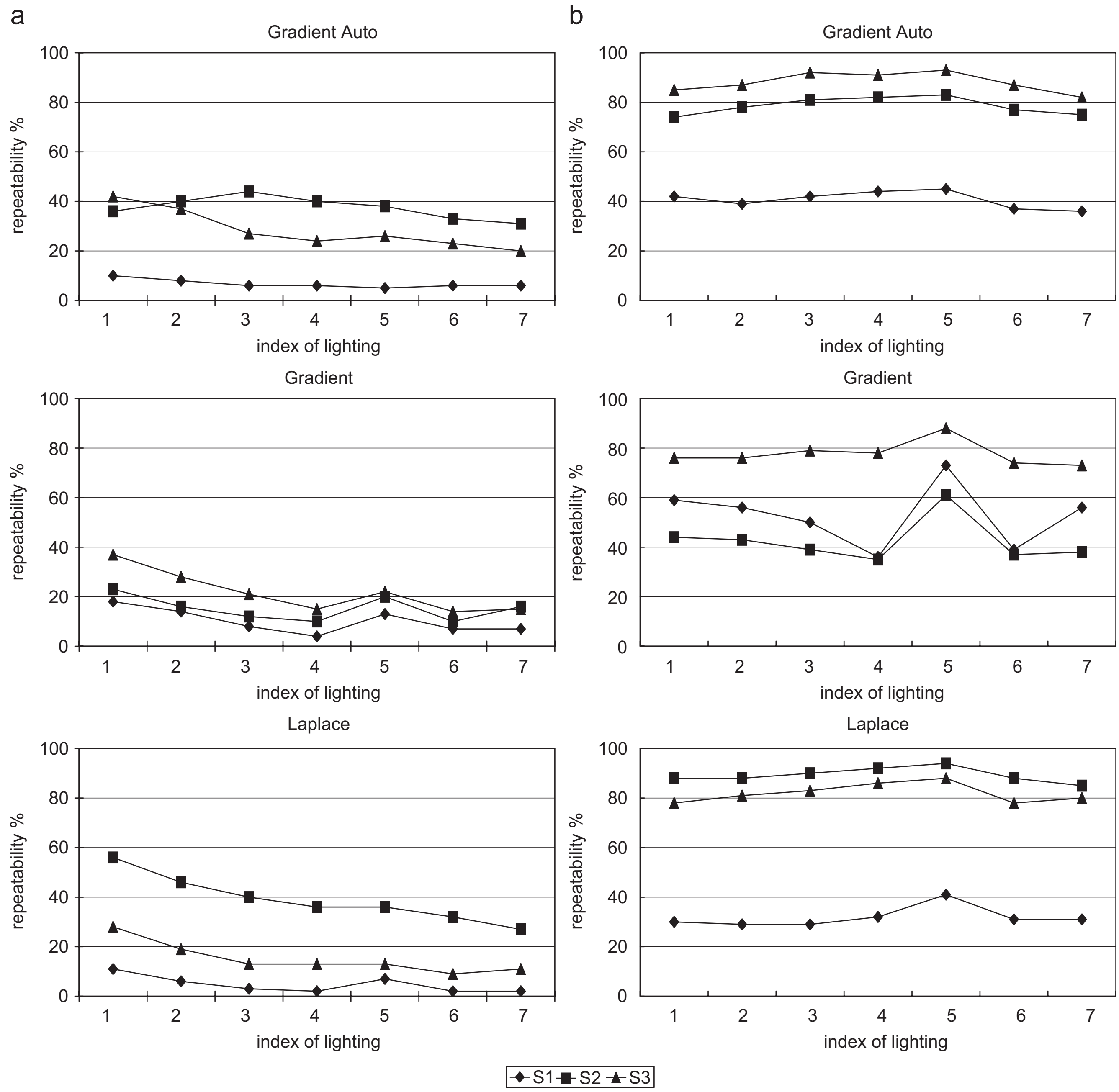

Fig. 14. ALOI: ε-repeatability rates of nine detectors across lighting conditions. (a) $\varepsilon_1 = 0.5$ and (b) (a) $\varepsilon_2 = 1.5$.

the repeatability on the face image. Fig. 17 gives us a more concrete view on the difference between weakly textured and richly textured images.

Now, we give a comparison of three detectors—Harris, DoG and GradientAuto/S3—on the sower image, testing their repeatability with respect to rotations. The results are shown in Fig. 18. We can observe that all three detectors are robust with respect to rotations on the sower image. It is also clear that the performance of three detectors is comparable to each other. This suggests that being applied to richly textured images, imbalance oriented selection does not show superiority over non-maximum suppression.

5. A new face recognition application: identity representability

Facial identity representation addresses the problem on how to select "good" facial instances for a gallery, as illustrated in Fig. 19.

Much previous work on face recognition used the neutral expressions as the gallery data, mostly motivated by the convenience of collecting such data. For example, the study in Ref. [26] is on the effect of the six important expressions as probe data, while using the neutral expression as gallery data. The appearance representation proposed in Ref. [26] is the weighted

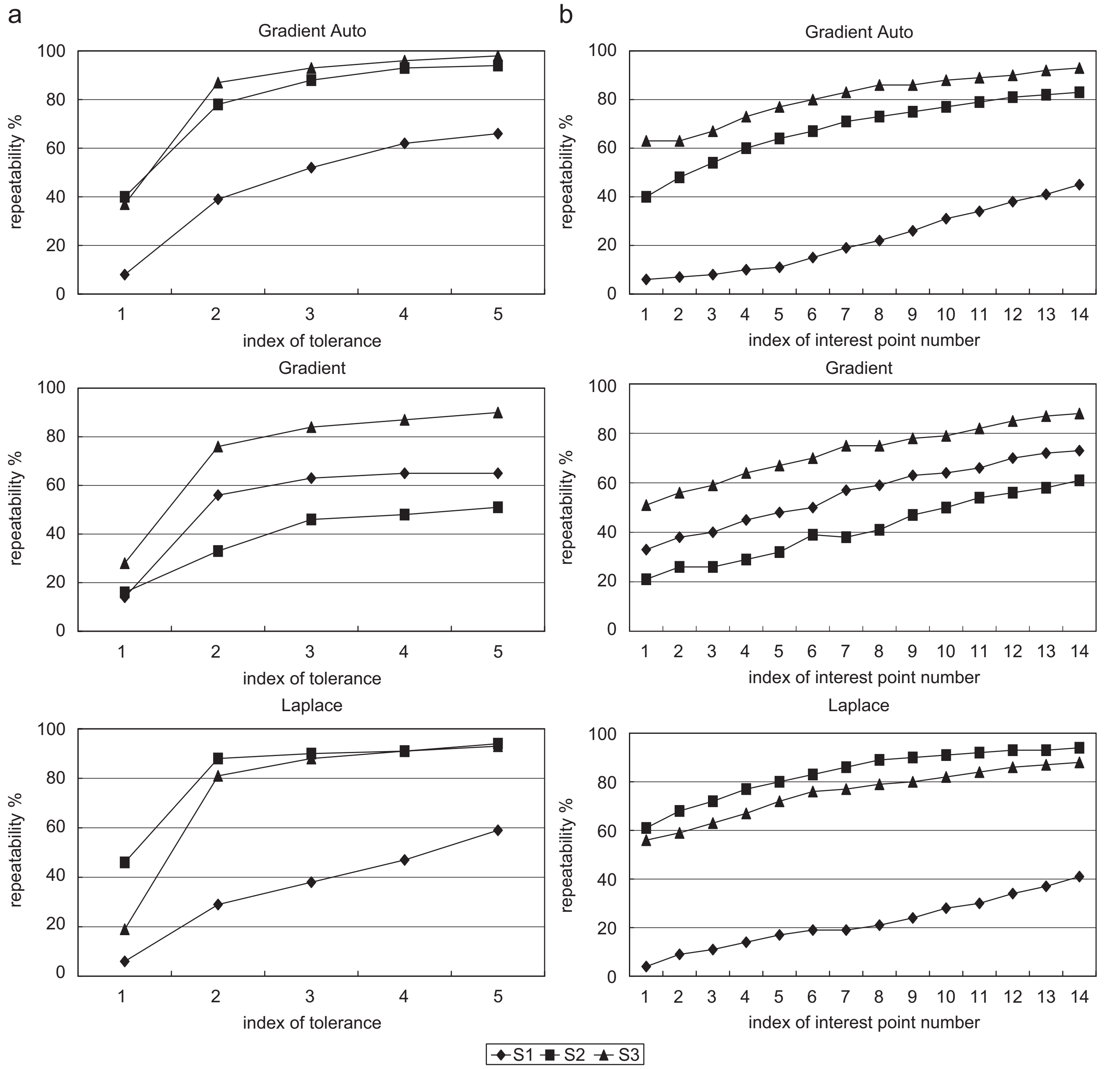

Fig. 15. ALOI: (a) ε-repeatability rates of nine detectors with respect to localization tolerances (from $\varepsilon_1 = 0.5$ to $\varepsilon_5 = 4.5$); and (b) $\varepsilon_2 = 1.5$-repeatability rates of nine detectors with respect to interest points numbers (from $n_1 = 140$ to $n_{14} = 400$).

global intensities via the optical flow technique. The authors of Ref. [28] presented a study on simultaneous face and facial recognition using facial expression decomposition, which did not address the identity representation problem either. The authors of Ref. [29] proposed an automatic feature localization for facial expression recognition using thermal images.

As a step towards the identity representation problem, the study in Ref. [30] is on the discriminant power of certain facial expressions, in terms of the Fisherface representation. The discriminant power is basically defined by the trace of $S_w^{-1}S_b$, where S_w and S_b are the within-class and between-class scatters of training expressions, respectively.

In this section, we present an evaluation of the identity representability of facial expressions based on spatial distributions of interest points. Our evaluation is currently focused on three facial expressions: *neutral, happiness*, and *anger*. The *identity representability* of an expression is measured by the recognition accuracy achieved by using its samples as the gallery data. Note that the identity representability of an expression can provide a quantified measurement of "goodness of the expression" that is selected for the gallery. Our experimental result will provide a rationale under the common utilization of neutral expression as gallery data in previous work.

5.1. Feature distributions

A *feature distribution* of an image is based on the number of occurrences of interest points in regular grids of an image plane. Fig. 20 shows the flow of generating a feature distribution for a simple input image. Given a facial appearance instance A, a feature distribution of A encodes its local and global appearance information: the number of interest points in each grid (implicitly) encodes a certain local appearance, and the alignment of all grids contributes to the global appearance information.

Fig. 16. Sower image [10].

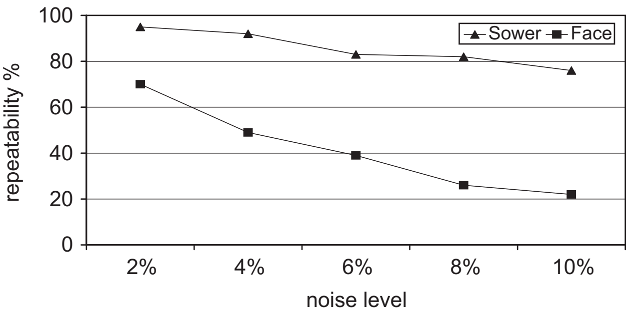

Fig. 17. Repeatability of Harris detector applied to Sower and Face images under different levels of noise.

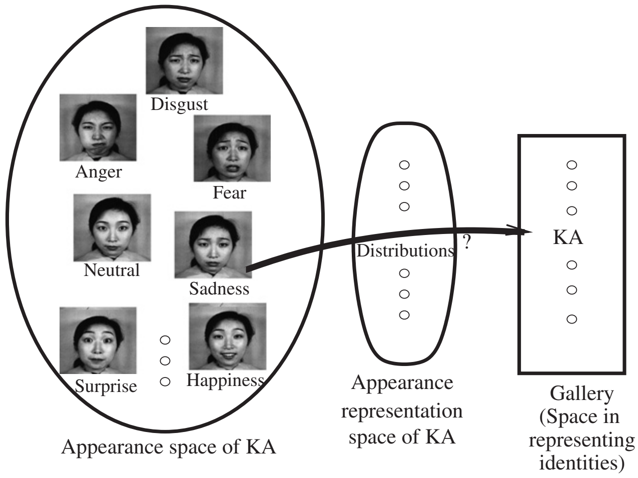

Fig. 19. Facial identity representation problem studies what facial instances, in certain appearance representations, are “good” to be selected for a gallery. (KA is one of ten facial identities in JAFFE [27].)

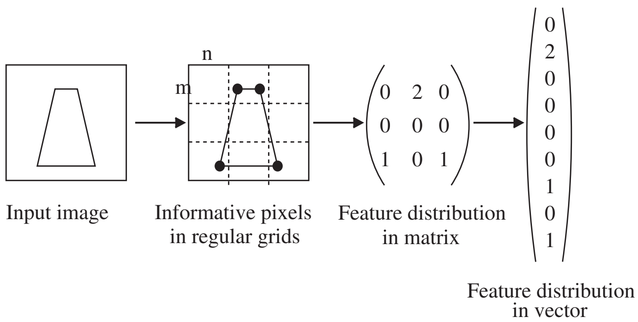

Fig. 20. The flow of generating a feature distribution for a simple input image.

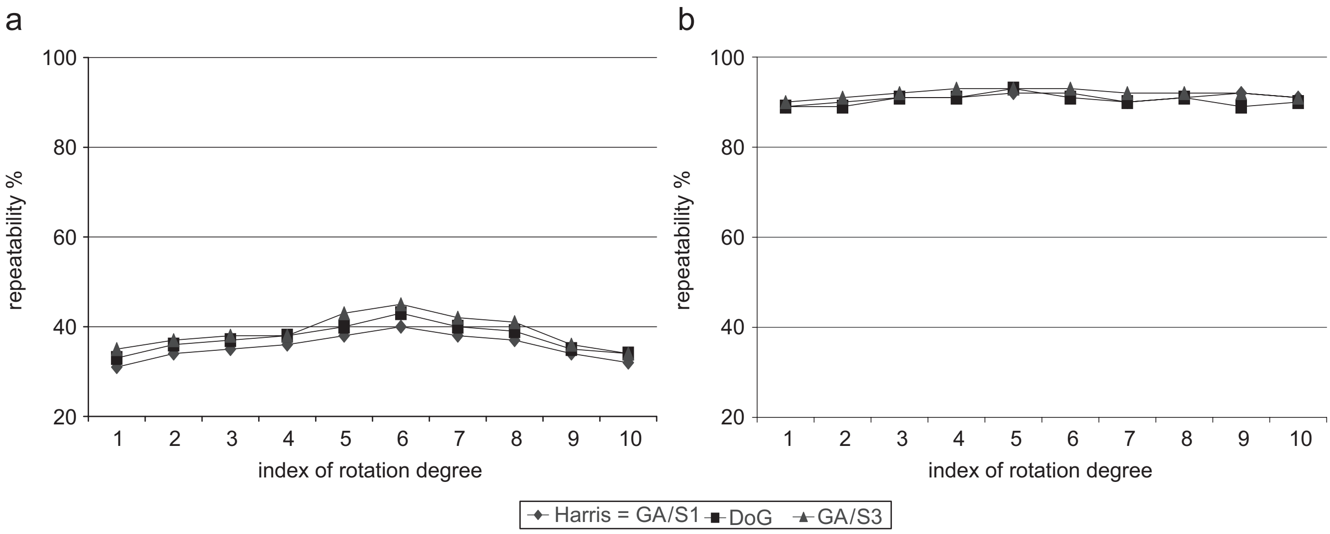

Fig. 18. Sower: ε-repeatability rate with respective to rotations: (a) $\varepsilon_1 = 0.5$ and (b) $\varepsilon_2 = 1.5$.

An appropriately large size of interest points collection P usually contains more complete information on local and global appearances. The grid size (m, n), however, should be small enough to characterize the local appearance information accurately. In the experimental evaluation, we use the following parameter settings: $|P| = 300$ and $(m, n) = (4, 8)$. The L^1-norm is used to measure the distance of two feature distributions. More details on the parameter settings can be found in our previous work [15].

5.2. Evaluation

The data sets used for the evaluation are JAFFE [27] and AR [31]. The JAFFE data set contains 10 identities (all females) [27]. The face regions in JAFFE images are well aligned, and our experiments use the raw images directly. To evaluate the identity representability of a specific expression, a single sample of the expression per identity is used as training and gallery data, and all others are used as probe data.

The subset of AR contains 126 identities. The face regions in AR images are not well aligned, and we manually aligned them. Fig. 21 shows three aligned AR images. A single sample in a specific expression per identity is used as training and gallery data, and all others are used as probe data. The nearest neighbor is used as the classifier. We use the following conventions in the rest of the paper: NE=neutral, HA=happiness, and AN=anger.

Fig. 21. AR images of facial expressions.

Fig. 22(a) and (b) shows the evaluation results on JAFFE and AR, respectively. We can observe that different feature distributions agree on the strongest identity representability of the *neutral* expression fairly consistently. There is an exception that is associated with the distribution generated by the GradientAuto/S1 detector. (Recall that S1 indicates non-maximum suppression.) We can also observe that different feature distributions agree on the weakest identity representability of the *anger* expression consistently.

In contrast to non-maximum suppression, the feature distributions generated by first-order imbalance achieve not only better consistency in the evaluation task but also higher recognition accuracy. Recall that non-maximum suppression can destroy local geometry (as illustrated in Fig. 1), and thus degrades the reliability of the feature distribution that it contributes.

6. Conclusions

In this paper, we propose imbalance oriented selections to detect interest points in weakly textured images. Without involving any suppression window, imbalance oriented selections can output interest points non-uniformly, i.e., it may detect a larger amount of points in certain regions than other regions, and thus provides a wise solution for the two issues of non-maximum suppression. The results from the test of repeatability across rotations and lighting conditions convince us the superiority of imbalance oriented selection over non-maximum suppression. We further present a new face recognition

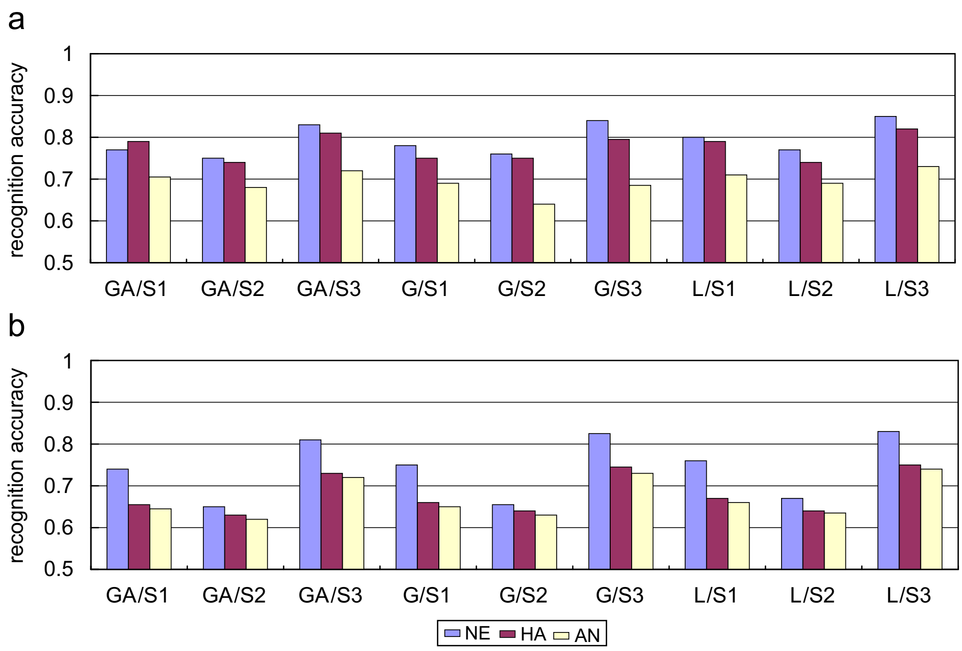

Fig. 22. Evaluation on: (a) JAFFE and (b) AR. GA, Gradient auto-correlation; G, Gradient; L, Laplace; S1, non-maximum suppression; S2, zero-order imbalance; and S3, first-order imbalance.

application where feature distributions are applied to evaluate identity representability of facial expressions.

Acknowledgments

The authors would like to thank the reviewers and the associate editor for their comments, which helped improve the paper significantly. The research of Q. Li and C. Kambhamettu was sponsored by the following grants: (i) Grant Number 2 P20 RR016472-06 under the INBRE program of the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and (ii) National Science Foundation (NSF) Office of Polar Programs (Grant number OPP0612105). The work of J. Ye was supported by National Science Foundation Grant IIS-0612069 and the Center for Evolutionary Functional Genomics of the Biodesign Institute at Arizona State University.

References

[1] C. Schmid, R. Mohr, Local grayvalue invariants for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 19 (5) (1997) 530–534.
[2] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[3] T. Deselaers, D. Keysers, H. Ney, Discriminative training for object recognition using image patches, in: Proceedings of the Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 157–162.
[4] C. Harris, M. Stephens, A combined corner and edge detector, in: Proceedings of the Fourth Alvey Vision Conference, Manchester, 1988, pp. 147–151.
[5] D. Reisfeld, H. Wolfson, Y. Yeshurun, Detection of interest points using symmetry, in: IEEE International Conference on Computer Vision, 1990, pp. 62–65.
[6] W. Forstner, A framework for low level feature extraction, in: European Conference on Computer Vision, 1994, pp. 383–394.
[7] H. Wang, M. Brady, Real-time corner detection algorithm for motion estimation, Image Vis. Comput. 13 (9) (1995) 695–703.
[8] S. Smith, J. Brady, SUSAN—a new approach to low level image processing, Int. J. Comput. Vis. 23 (1) (1997) 45–78.
[9] E. Loupias, N. Sebe, Wavelet-based salient points for image retrieval, in: RR 99.11, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon, November 1999.
[10] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest point detectors, Int. J. Comput. Vis. 37 (2) (2000) 151–172.
[11] K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors, Int. J. Comput. Vis. 60 (1) (2004) 63–86.
[12] F. Faille, A fast method to improve the stability of interest point detection under illumination changes, in: International Conference on Image Processing, 2004, pp. 2673–2676.
[13] B. Platel, E. Balmachnova, L. Florack, F. Kanters, B.M. ter Haar Romeny, Using top-points as interest points for image matching, in: Deep Structure, Singularities, and Computer Vision, 2005, pp. 211–222.
[14] C.S. Kenney, M. Zuliani, B.S. Manjunath, An axiomatic approach to corner detection, in: International Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 191–197.
[15] Q. Li, J. Ye, C. Kambhamettu, Spatial interest pixel (SIPs): useful low-level features of visual media data, in: IEEE International Conference on Data Mining, 2003, pp. 163–170.
[16] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: IJCAI, 1981, pp. 674–679.
[17] K. Mikolajczyk, C. Schmid, Indexing based on scale invariant interest points, in: IEEE International Conference on Computer Vision, vol. I, Vancouver, Canada, 2001, pp. 525–531.
[18] G. Olague, B. Hernandez, A new accurate and flexible model-based multi-corner detector for measurement and recognition, Pattern Recognition Lett. 26 (1) (2005) 27–41.
[19] L. Trujillo, G. Olague, Using evolution to learn how to perform interest point detection, in: International Conference on Pattern Recognition, August 20–24, 2006, Hong Kong, China, 2006, 211–214.
[20] L. Trujillo, G. Olague, Synthesis of interest point detectors through genetic programming, in: Genetic and Evolutionary Computation Conference, July 8–12, 2006, Seattle, WA, USA, 2006, pp. 887–894.
[21] M. Brown, R. Szeliski, S. Winder, Multi-image matching using multi-scale oriented patches, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 510–517.
[22] H. Sahbi, N. Boujemaa, Robust face recognition using dynamic space warping, in: ECCV's 2002 Workshop on Biometric Authentication, 2002.
[23] R. Paredes, J.C. Perez, A. Juan, E. Vidal, Face recognition using local representations and a direct voting scheme, in: Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis, 2001.
[24] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression (PIE) database, in: Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 2002, pp. 53–58.
[25] J.M. Geusebroek, G.J. Burghouts, A.W.M. Smeulders, The Amsterdam library of object images, Int. J. Comput. Vis. 61 (1) (2005) 103–112.
[26] A.M. Martinez, Recognizing expression variant faces from a single sample image per class, in: IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. 353–358.
[27] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Trans. Pattern Anal. Mach. Intell. 21 (12) (1999) 1357–1362.
[28] H. Wang, N. Ahuja, Facial expression decomposition, in: International Conference on Computer Vision, 2003, pp. 958–965.
[29] L. Trujillo, G. Olague, R. Hammoud, B. Hernandez, Automatic feature localization in thermal images for facial expression recognition, in: Second Joint IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum, OTCBVS 2005, In Conjunction with CVPR 2005, vol. 3, 2005, p. 14.
[30] Y. Yacoob, L. Davis, Smiling faces are better for face recognition, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 59–64.
[31] A.M. Martinez, R. Benavente, The AR face database, Technical Report CVC, Technical Report No. 24, 1998.

About the Author—QI LI is an Assistant Professor of the Department of Computer Science at Western Kentucky University. He received his Ph.D. in Computer Science from University of Delaware in 2006. His current research interest include pattern recognition, computer vision, machine learning, and bioinformatics.

About the Author—JIEPING YE is an Assistant Professor of the Department of Computer Science and Engineering at the Arizona State University. He received his Ph.D. in Computer Science from University of Minnesota, Twin Cities in 2005. His research interests lie in machine learning, data mining, and bioinformatics. In 2004, his paper on generalized low rank approximations of matrices won the outstanding student paper award at the 21st International Conference on Machine Learning. He is a member of IEEE.

About the Author—CHANDRA KAMBHAMETTU received his M.S. and Ph.D. degrees in Computer Science and Engineering from the University of South Florida in 1991 and 1994, respectively. From 1994 to 1996, he was a research scientist at NASA-Goddard, where he received the 1995 Excellence in Research Award from the Universities Space Research Association (USRA). From 1996 to 1997, he was a visiting faculty at the Department of Computer and Information Sciences, University of Delaware. He was an Assistant Professor from 1997 to 2003, and is an Associate Professor from 2003–present in

the same department where he leads the Video/Image Modeling and Synthesis (VIMS) group. Dr. Kambhametu received NSF CAREER award in 2000. His research interests include computer vision, computer graphics, biomedical image analysis, bioinformatics and remote sensing. Dr. Kambhamettu is best known for his work in nonrigid motion analysis of deformable bodies. He published over 90 papers in this area. He currently serves as special issue guest editor on this topic in Image and Vision Computing journal. He is Associate Editor for journals Pattern Recognition, and IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).