# Image Feature Extraction Using Gradient Local Auto-Correlations

Takumi Kobayashi and Nobuyuki Otsu

National Institute of Advanced Industrial Science and Technology,
1-1-1 Umezono, Tsukuba, Japan
{takumi.kobayashi,otsu.n}@aist.go.jp

**Abstract.** In this paper, we propose a method for extracting image features which utilizes 2 nd order statistics, i.e., spatial and orientational auto-correlations of local gradients. It enables us to extract richer information from images and to obtain more discriminative power than standard histogram based methods. The image gradients are sparsely described in terms of magnitude and orientation. In addition, normal vectors on the image surface are derived from the gradients and these could also be utilized instead of the gradients. From a geometrical viewpoint, the method extracts information about not only the gradients but also the curvatures of the image surface. Experimental results for pedestrian detection and image patch matching demonstrate the effectiveness of the proposed method compared with other methods, such as HOG and SIFT.

## 1 Introduction

Extracting features from an image is a fundamental procedure for various tasks, e.g., face or human detection [1,2], image patch matching [3], object recognition [4] and image retrieval [5]. It is important to extract characteristics of target objects and textures with retaining robustness to irrelevant variations resulting from environmental changes, such as changes in illumination or target position. Strictly speaking, we can identify two types of image features by focusing on image alignments: a shift-invariant type and a local image descriptor type.

The former type needs object regions not to be aligned and thus has the property of shift-invariance for the target objects. Fourier transformation and histogram based methods are traditionally applied to this type. This property of shift-invariance is particularly favorable for the task of object recognition, since it can then be carried out irrespective of the target position. However, it is difficult to obtain sufficient discriminative power for this type of features.

The latter type assumes aligned object regions and it is often dealt with in terms of a local image descriptor [3], which takes advantage of spatial alignment in the image region. The features of this type have been successfully developed and they play important roles, especially for image patch matching problems. These features include small patch [6], Shape Context [7], self similarity [8] and image gradients [9]. A comprehensive survey of local image descriptors is given in [3]. These local descriptors have been recently utilized in bag-of-feature frameworks which work particularly well for object recognition [10,4,11]. On the other hand, the shift-invariant features mentioned above

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part I, LNCS 5302, pp. 346–358, 2008.

can be naturally applied to local descriptors by simply dividing regions into several subregions (spatial binning), as in SIFT [12] and HOG [13].

In this paper, we propose a method for extracting shift-invariant image features which can also be applied as local descriptors. It extracts richer information, i.e., 2 nd order statistics of gradients, and thus obtains more discriminative power than standard histogram based methods. The proposed method is based on spatial and orientational auto-correlations of local image gradients: Gradient Local Auto-Correlation (GLAC). In GLAC, the image gradients are described sparsely in terms of their magnitude and orientation. Furthermore, the gradients can be extended to normal vectors on the image surface, which can be utilized for Normal Local Auto-Correlation (NLAC). We applied the proposed methods, GLAC and NLAC, as local image descriptors to two tasks: human detection and image patch matching. The experimental results demonstrate their effectiveness compared with other methods, including SIFT and HOG.

## 2 Related Work

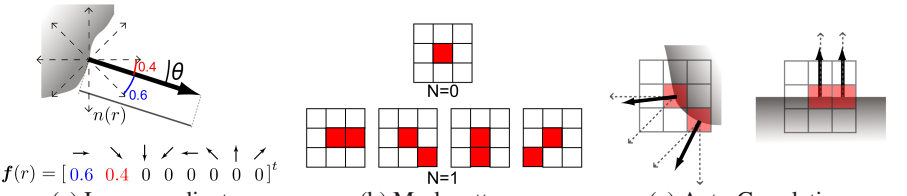
We mention here only closely related work. SIFT [12] and HOG [13] are some of the most successful features based on histograms of gradient orientations weighted by gradient magnitudes. These two methods slightly differ in the type of spatial bins that they use; HOG has a more sophisticated way of binning. The concept of correlation has also been adopted in self similarity [8], in which extracted edges in Shape Context [7] are substituted with cross-correlation values between local patches at a reference position and its local neighborhoods. Our work is most closely related to ECM [14] which utilizes joint histograms of orientations of gradient pairs. Differences in the details are described in Sec.3.2.

## 3 Gradient Local Auto-Correlations

In this section, we describe the details of the proposed method, Gradient Local Auto-Correlations (GLAC). It can be interpreted as a natural extension of HOG or SIFT from 1 st order statistics (i.e., histograms) to 2 nd order statistics (i.e., auto-correlations). In GLAC, image gradients are sparsely described in terms of their magnitudes and orientations. The proposed formulation extends naturally from Higher-order Local Auto-Correlation (HLAC) [15] of pixel values so as to deal with gradients as well. Therefore, GLAC inherits the desirable properties of HLAC for recognition: *shift-invariance* and additivity.

### 3.1 Definition of GLAC

Let $I$ be an image region and $\boldsymbol{r}=(x,y)^t$ be a position vector in $I$. The image gradient $(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})^t$ at each pixel can be rewritten in terms of the magnitude $n=\sqrt{\frac{\partial I}{\partial x}^2+\frac{\partial I}{\partial y}^2}$ and the orientation angle $\theta=\arctan(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})$. As shown in Fig. 1(a), the orientation $\theta$ is coded into $D$ orientation bins by voting weights to the nearest bins, and is described as a sparse vector $\boldsymbol{f}(\in \boldsymbol{R}^D)$, called the *gradient orientation vector* (in short, G-O vector).




(a) Image gradient (b) Mask patterns (c) Auto-Correlation

**Fig. 1.** Image gradients are described by the G-O vectors, together with the gradient magnitudes (a). Then, by applying mask patterns (b), auto-correlations of G-O vectors are calculated, weighted by the gradient magnitudes (c).

It is important that the image gradients are represented in terms of such quantized and sparse descriptors.

By using the G-O vector $\boldsymbol{f}$ and the gradient magnitude $n$, the $N^{\text{th}}$order auto-correlation function of gradients in local neighbors is defined as follows:

$$R(d_0, \cdots, d_N, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_N) = \int_I w\big[n(\boldsymbol{r}), n(\boldsymbol{r}+\boldsymbol{a}_1), \cdots, n(\boldsymbol{r}+\boldsymbol{a}_N)\big] f_{d_0}(\boldsymbol{r}) f_{d_1}(\boldsymbol{r}+\boldsymbol{a}_1) \cdots f_{d_N}(\boldsymbol{r}+\boldsymbol{a}_N) d\boldsymbol{r}, \tag{1}$$

where $\boldsymbol{a}_i$ are displacement vectors from the reference point $\boldsymbol{r}$, $f_d$ is the $d$-th element of $\boldsymbol{f}$ and $w$ is a (scalar) weighting function, e.g., $\min$. Displacement vectors are limited to local neighbors because local gradients are supposed to be highly correlated.

Eq.(1) contains two kinds of correlations of gradients: *spatial* correlations derived from displacement vectors $\boldsymbol{a}_i$ and *orientational* correlations derived from the products of the element values $f_{d_i}$. We do not correlate image gradients themselves but G-O vectors which are quantized and represented sparsely. This is due to the empirical fact that, in HLAC [15], the auto-correlations of binary values, i.e., quantized data, are better for establishing recognition than those of the pixel values themselves. The function $w$ composed of magnitudes $n$ functions as the weights of the auto-correlation.

In practice, Eq.(1) can take so many forms by varying the parameters $N, \boldsymbol{a}_i$, and the weight $w$. In this paper, these are restricted to vary as follows: $N \in \{0, 1\}$, $a_{1x,y} \in \{\pm\Delta r, 0\}$, and $w(\cdot) \equiv \min(\cdot)$. The order of auto-correlation, $N$, is low, which enables extraction of sufficient geometric characteristics together with local displacements $\boldsymbol{a}_i$. The displacement intervals are the same in both horizontal and vertical directions due to isotropy of the image. We adopt $\min$ for $w$ in order to possibly suppress the effect of isolated noise on surrounding auto-correlations. Thus, the practical formulation of GLAC is given by

$$\textbf{0}^{\textbf{th}}\textbf{order} \quad R_{N=0}(d_0) = \sum_{\boldsymbol{r} \in I} n(\boldsymbol{r}) f_{d_0}(\boldsymbol{r}) \tag{2}$$

$$\textbf{1}^{\textbf{st}}\textbf{order} \quad R_{N=1}(d_0, d_1, \boldsymbol{a}_1) = \sum_{\boldsymbol{r} \in I} \min\big[n(\boldsymbol{r}), n(\boldsymbol{r}+\boldsymbol{a}_1)\big] f_{d_0}(\boldsymbol{r}) f_{d_1}(\boldsymbol{r}+\boldsymbol{a}_1).$$

The configuration patterns of $(\boldsymbol{r}, \boldsymbol{r}+\boldsymbol{a}_1)$, i.e., the spatial auto-correlation patterns, are shown in Fig. 1(b). It should be noted that we obtain only four independent patterns

**Algorithm 1.** GLAC computation

**Preprocessing:** The G-O vector $\boldsymbol{f}$ and the gradient magnitude $n$ are calculated from image gradients.
**0 th order:** At each pixel $\boldsymbol{r}$, summation in Eq.(2) is applied to only *two* non-zero elements of $\boldsymbol{f}$ with weight $n$.
**1 st order:** At each pixel $\boldsymbol{r}$, for each mask pattern (Fig. 1(b)), summation of products in Eq.(2) are applied to non-zero elements of $\boldsymbol{f}(\boldsymbol{r})$ and $\boldsymbol{f}(\boldsymbol{r}+\boldsymbol{a}_1)$ with weight of $\min[n(\boldsymbol{r}), n(\boldsymbol{r}+\boldsymbol{a}_1)]$. This takes only *four* times operations of multiplication.

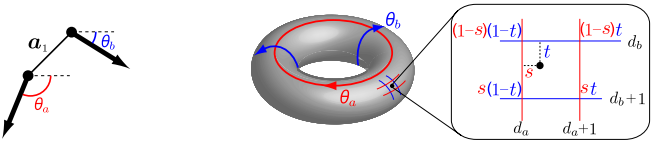
for 1 st order GLAC by eliminating duplicates which arise from shifts. For the 1 st order, the element values of G-O vector pairs determined by the mask patterns are multiplied and summed over the image (Fig. 1(c)). Although GLAC has high dimensionality ($D + 4D^2$), the computational cost is not large due to the sparseness of $\boldsymbol{f}$ (see Algorithm 1). Moreover, the computational cost of GLAC is invariant with respect to the number of orientation bins, $D$, since the sparseness of $\boldsymbol{f}$ is invariant with respect to $D$. In the case of calculating features in many sub-regions of an image, we can apply a method similar to the *integral image* approach [1], which is particularly effective for the object detection problem, e.g., face or pedestrian detection.

### 3.2 Interpretation

**Histogram.** While 0 th order GLAC simply corresponds to a histogram of gradient orientations used in SIFT [12] and HOG [13], 1 st order can be interpreted as a joint histogram of orientation pairs. Now, we consider the joint distribution of orientation pairs of local gradients, taking into account the fact that the orientation angles are periodic in $[0, 2\pi)$. Given a certain displacement vector $\boldsymbol{a}_1$ which determines the local pairs (Fig. 2(a)), the orientation pairs are jointly distributed on the torus manifold defined by the paired angles (Fig. 2(b)). The 1 st order GLAC corresponds to the joint histogram calculated by quantizing the distribution into $D \times D$ bins on the torus with *bilinear* weighting (Fig. 2(b)). This joint histogram weighted by $w$ forms 2 nd order statistics naturally extended from the histogram of orientations (1 st order statistics). From this perspective, the 0 th order GLAC is a marginal histogram of the 1 st order. This suggests that the 0 th order components are not independent of the 1 st order and may be redundant, which is verified by experiments (see Sec.5).

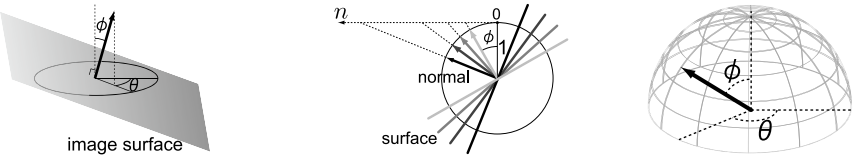
ECM [14] also utilizes a joint histogram of orientation pairs, but it is a special case of GLAC: $w \equiv 1$ (no weighting) and the G-O vector consists of binary values (0 or 1) in ECM. It suffers from boundary effects of the magnitude and the orientation of image gradients. Moreover, the displacement vectors are not specified in ECM, whereas they are determined according to the auto-correlation scheme in this paper.

**Geometry.** The 1 st order GLAC characterizes curvatures of image contours. The curvatures are quantized and patterned by the combinations of the orientations of local gradient pairs in the mask pattern as shown in Fig. 1(c). GLAC extracts image features in terms of gradients and curvatures which are fundamental properties of the image contours. In GLAC, the curvatures are distinguished by rotation. Rotational invariance,




(a) Gradient pair (b) Joint histogram on torus manifold

**Fig. 2.** GLAC is a joint histogram of paired angles on the torus manifold (b) determined by a displacement vector (a)




(a) Normal vector (b) Steepness of image surface (c) Hemisphere

**Fig. 3.** Normal vector to image surface

however, can be rendered by simply summing up the component values associated with curvature patterns which are matched by rotation.

Next, we consider the image surface defined by pixel values in 3-D space denoted as $\boldsymbol{z} = (x, y, I(x, y))^t$. The normal vector to the surface is calculated as follows:

$$\frac{\partial \boldsymbol{z}}{\partial x} \times \frac{\partial \boldsymbol{z}}{\partial y} = \left(-\frac{\partial I}{\partial x}, -\frac{\partial I}{\partial y}, 1\right)^t, \quad \phi = \arctan \sqrt{\left[\frac{\partial I(x,y)}{\partial x}\right]^2 + \left[\frac{\partial I(x,y)}{\partial y}\right]^2}, \qquad (3)$$

where $\phi$ is the angle of elevation (Fig. 3(a)). Thus, the gradient magnitude $n$ determines the steepness of the local surface (Fig. 3(b)). The weight $w$ controlled by $n$ corresponds to the magnitude of the curvature on the image surface in 3-D space and, consequently, GLAC focuses on principal curvatures by means of weightings.

## 4 Normal Local Auto-Correlations

The normal vectors (Fig. 3) can be employed instead of the gradients described in Sec.3.1. The normals characterize the image surface in 3-D space while the gradients do the same for the image contours in a 2-D image plane. Thus, by using normals, Normal Local Auto-Correlation (NLAC) can be developed to extract the detailed features of the image surface, in a manner similar to GLAC.

### 4.1 Normal Orientation Vector

As shown in Fig. 3(a) and Eq.(3), a normal vector is characterized by the orientation $\theta$ in the x-y plane and the angle of elevation $\phi$. The normal can be coded by *bilinear* weighting on the hemisphere composed of two angles $\theta, \phi$ (Fig. 3(c)) and then the *normal orientation vector* (N-O vector) $\boldsymbol{g}$ can be defined in a manner similar to the G-O vector in Sec.3.1.

Here, the problem is how to define the scale of pixel values $I(x, y)$ in Eq.(3). The scale of $\partial I(x, y)$ (the pixel value domain), which is arbitrarily defined by users, e.g., $[0, 1]$ or $[0, 255]$, is intrinsically different from that of $\partial x, \partial y$ (the pixel location domain). Let a pixel value be $I_o$ in certain scale, e.g., $[0, 1]$. The elevation angle $\phi$ in Eq.(3) can be rewritten as

$$\phi = \arctan\left( k\sqrt{\left[\frac{\partial I_o(x,y)}{\partial x}\right]^2 + \left[\frac{\partial I_o(x,y)}{\partial y}\right]^2} \right) = \arctan(kn) \tag{4}$$

where $k$ is a scaling factor. The problem is how to determine $k$ appropriately so as to be consistent with $\partial x, \partial y$.

The scaling $k$ determines the distribution of normals on the hemisphere: if $k \to 0$, the normals would be concentrated near the zenith and, contrarily, if $k \to \infty$ they would be located only around the periphery. From the viewpoint that the normals are coded into equally spaced bins on the hemisphere and are described as the N-O vector, the scaling $k$ can be determined so that the distribution of the normals is uniform along $\phi$ in order to make all bins on the hemisphere useful. In this case, the distribution is transformed by the function $\arctan(kn)$. In terms of histogram equalization [16], $\arctan(kn)$ is required to be similar to the probability distribution function of gradient magnitude $n$ in order to make uniform distribution on $\phi$. Thus, the scaling $k$ is determined as

$$k = \arg\min_{k,l} \left| P(n) - l \arctan(kn) \right|^2 \tag{5}$$

where $P$ is the probability distribution function of $n$ and $l$ is introduced so as to fit the ranges of $P$ and $\arctan$, which does not affect the distribution.

### 4.2 Definition of NLAC

By using the N-O vector $\boldsymbol{g}$, NLAC can be computed as

$$R_{N=0}(d_0) = \sum_{\boldsymbol{r} \in I} g_{d_0}(\boldsymbol{r}), \quad R_{N=1}(d_0, d_1, \boldsymbol{a}_1) = \sum_{\boldsymbol{r} \in I} g_{d_0}(\boldsymbol{r}) g_{d_1}(\boldsymbol{r} + \boldsymbol{a}_1) d\boldsymbol{r}. \tag{6}$$

This does not include weighting whereas the weight derived from the gradient magnitude $n$ is utilized in GLAC (Eq.(1)). This is because the N-O vector $\boldsymbol{g}$ already contains information about the magnitude in the angle of elevation $\phi$. The computational cost of NLAC is small due to the sparseness of $\boldsymbol{g}$ as well as that of GLAC.

From a geometrical viewpoint, NLAC is a histogram of patterns of curvatures on the image surface. Although, in GLAC, the principal curvatures on the surface are highly weighted, all patterns of the curvatures can be captured in NLAC regardless of the magnitudes of the curvatures. For example, a curvature which includes a zero gradient, e.g., Fig. 4(a), is disregarded in GLAC, but it is counted in NLAC. However, the count of a flat curvature, e.g., Fig. 4(b), is closely related to the area size of the object. This is a square function of the target scale, which reduces robustness to the scale. Therefore, when the target scale is not normalized, we disregard curvature patterns arising from flatness which are related to only one element of $\boldsymbol{g}$ associated with the zenith bin on the hemisphere. The number of disregarded patterns is 1 ($0^{\text{th}}$order) + 4 ($1^{\text{st}}$order) = 5.

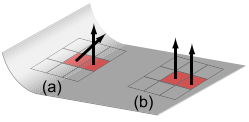
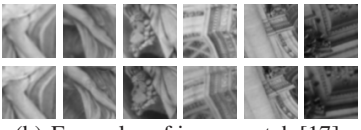



**Fig. 4.** NLAC can capture various patterns of curvatures, even those containing zero gradients: (a) foot of hill and (b) flatness. The curvature pattern of (b) would be disregarded.

(a) Examples of person image [13]



(b) Examples of image patch [17]

**Fig. 5.** Images in datasets

## 5 Experimental Results

We apply the proposed methods to two kinds of task: human detection [13] and image patch matching [17] in order to compare the performances with those of HOG [13] and SIFT [12] which have been some of the most successful methods for these tasks.

### 5.1 Human Detection

In this experiment, the extracted features are classified by using the linear SVM [18]. The proposed methods were tested on the INRIA person dataset (Fig. 5(a)), details of which are in [13]. We selected 2416 person and 12180 person-free images (64×128) for training, and 1132 person and 13590 person-free images for testing. For quantifying and comparing detectors, we plotted Receiver Operating Characteristics (ROC) curves by calculating False Positive (FP) and True Positive (TP) Rates.

Although GLAC and NLAC are completely shift-invariant, the detection problem does not require this property due to roughly aligned person images arising from shifting the detection window in the image. Thus, for accuracy comparisons, the image region is divided into sub-regions (blocks), e.g., 4×4 blocks, and the GLAC/NLAC features extracted in these blocks are integrated into a final feature vector in the same manner as SIFT [12]. Spatial binning reduces shift-invariance but increases discriminative power as shown in the next.

**Comparison to the other methods.** First, we compare overall performances of the proposed methods with those of the other methods: HOG [13], Steerable Filter [19,17] and Steerable Filter Local Auto-Correlation (SLAC). HOG, for which the parameter settings are those of [13], has produced the best performance for this database. The Steerable Filter feature consists of the rectified response values of fourth order derivative filters [19], and this method has worked well in image patch matching [17]. SLAC is newly constructed here by using the Steerable Filter feature vector instead of $\boldsymbol{g}$ in Eq.(6). In the Steerable Filter and SLAC approaches, spatial binning is also applied. The

performance results are shown in Fig. 8(a). Both the GLAC and NLAC methods outperform the other methods including HOG. The performance of NLAC is lower than that of GLAC even when utilizing the same spatial bins, and these methods are compared in the last part of this section. GLAC with $4\times5$ blocks has a higher dimensionality than GLAC with $3\times4$ blocks, but results in further improvement. When $3\times4$ spatial bins are utilized, the dimension of GLAC features is almost the same as that of HOG. Note that the number of these spatial bins is significantly smaller than for HOG and thus larger spatial perturbations can be allowed. The performance of SLAC is a great improvement on that of Steerable Filter, but it is inferior to GLAC. This is because the G-O vector is much sparser than the Steerable Filter vector. As described before, auto-correlations work particularly well for sparse data.

**Performance Study.** Next, focusing on GLAC, we give details of the parameter settings and their effects on performance. We refer to the baseline parameter settings as: 1) the Roberts gradient filter; 2) 9 orientation bins in 360 degrees; 3) a spatial interval $\Delta r = 1$; 4) a weighting $w(\cdot) \equiv \min(\cdot)$; 5) only 1 $^{\text{st}}$order auto-correlation; 6) block-wise L2-Hys normalization; 7) $3\times4$ spatial blocks, which are the same as in Fig. 8(a).

**[Gradient]** Gradient computation is the first processing step that may affect the final performance. We applied three types of filters: Roberts, Sobel and one-dimensional derivatives ($[-1, 0, 1]$). The Roberts filter, which is the most compact, is most effective, whereas the smoothed Sobel filter is least effective (Fig. 8(b)). As shown in [13], smoothing the images results in reduced performance.

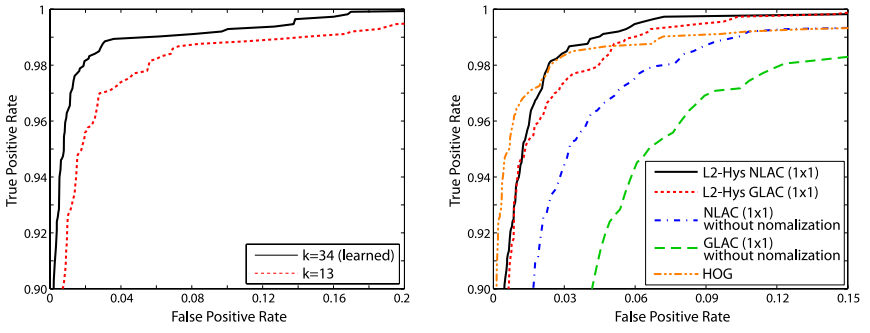
**[Orientation bins]** Orientation bins are evenly spaced over $[0^\circ, 180^\circ)$ (unsigned gradients) or $[0^\circ, 360^\circ)$ (signed gradients). Fig. 8(c) shows that finer binning increases performance. Contrary to the results in [13], the signed gradient works even better than the unsigned gradient. For auto-correlations of orientations, signed gradients seem to be preferable. The recent results of object recognition using HOG have also shown a similar tendency [11].

**[Spatial interval]** The only parameter in auto-correlations is the spatial interval $\Delta r$ which is closely related to the scale of the objects to be recognized. As shown in Fig. 8(d), small interval values, really *local* auto-correlations, work well with the compactness of the Roberts filter.

**[Weighting]** The weight $w$ in Eq.(1) is qualitatively defined as $\min$, taking the perspective of noise reduction. It is quantitatively compared with $\max$ and product ($\prod n$) in Fig. 8(e). As expected, $\min$ is the best, due to the noise reduction effect.

**[Correlation order]** The composition of GLAC can be varied as follows: only 1 $^{\text{st}}$order, both 0 $^{\text{th}}$and 1 $^{\text{st}}$order, and only 0 $^{\text{th}}$order. Fig. 8(f) shows that the addition of the 0 $^{\text{th}}$order to the 1 $^{\text{st}}$order has no, even worse, effect on performance. Thus, the 0 $^{\text{th}}$order components seem to be redundant, as suggested in Sec.3.2.

**[Normalization]** We adopted two types of normalization: L2 and L2-Hys. L2 refers to normalization by L2-norm and L2-Hys means clipping component values after L2 as in [12]. These normalizations are applied either to whole feature vector or block-wise. Fig. 8(g) shows that L2-Hys outperforms L2 while block-wise normalization is better




(a) Scaling factor (b) Comparison without spatial binning

**Fig. 6.** The detection performances of NLAC with various settings. Details are in the text.

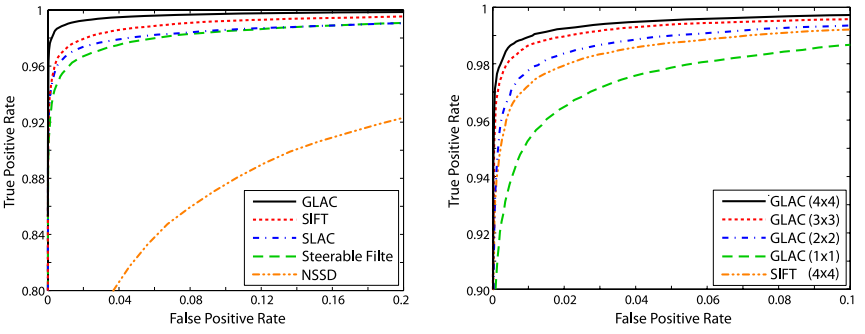
than whole normalization. In summary, block-wise L2-Hys normalization is the best, and performance is greatly improved, compared to performance without normalization.

**[Spatial bins]** Due to the dimensionality of GLAC, we applied somewhat coarser spatial binning, equally spaced over an image (64×128) as in [12]. As shown in Fig. 8(h), binning finer than 3×4 results in sufficiently good performance, and, in particular, 4×5 binning is most effective. Spatial binning results in greatly improved performance, compared to performance without spatial binning (1×1).

**NLAC.** In NLAC, the scaling factor $k$ in Eq.(4) is learned from the MIT pedestrian dataset [20]; $k$=34. Fig. 6(a) shows that the learned value of $k$ is appropriate and effective compared with a randomly chosen value of $k$=13. In Fig. 8(a), NLAC of 3×4 blocks outperforms HOG but it is inferior to GLAC. On the contrary, for no spatial binning (1×1) in Fig. 6(b), the performance of NLAC with L2-Hys is superior to that of GLAC. The effect of normalization (L2-Hys) on performance is greater for GLAC than for NLAC, by comparison of the results without normalization. In NLAC, the gradient magnitude $n$ is already transformed by $\arctan$ in Eq.(4) at each pixel, which reduces the effect of L2-Hys normalization as a nonlinear operation. It is noteworthy that, even when other processes (spatial binning and normalization) are not applied, NLAC gives high performance with retaining the favorable properties of shift-invariance and additivity. In summary, GLAC is better suited to local descriptors and NLAC is better suited to shift-invariant features.

### 5.2 Image Patch Matching

We applied the proposed method (GLAC) to local image descriptors for image patch matching on the database of image patches [17] (Fig. 5(b)). This database contains matched image patches (64×64) collected by using SIFT detector and descriptor [12] and further 3D point estimation from tourist photographs of Trevi Fountain, Yosemite Valley and Notre Dame. See [17] for details of the database. We followed the procedure in [17] for training and testing: 10,000 matched pairs and 10,000 non-matched



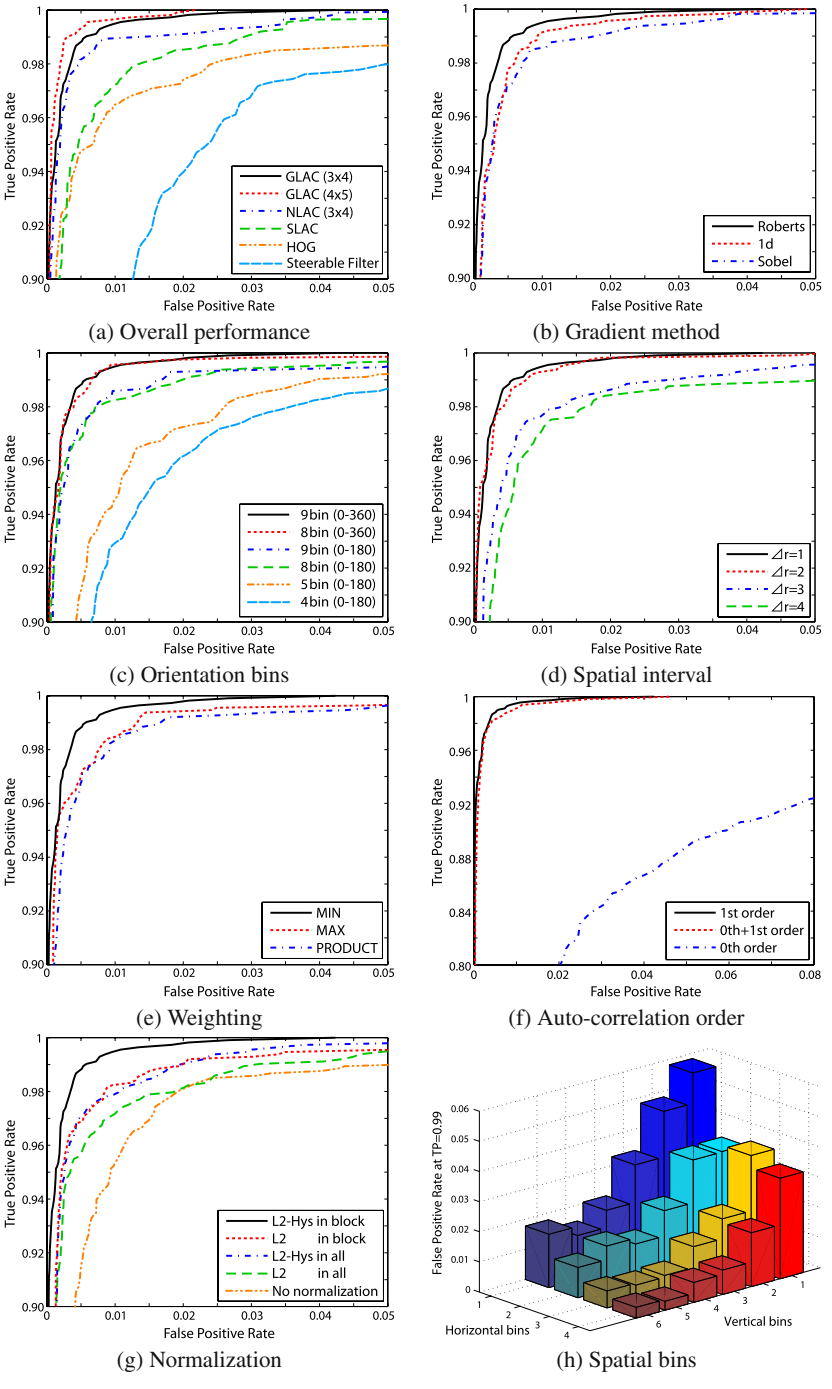

(a) Overall comparison (b) Spatial binning

**Fig. 7.** The results of image patch matching. Details are in the text.

pairs were randomly sampled from the Trevi and Yosemite datasets in order to learn parameters for the local image descriptors. For testing, 50,000 matched and 50,000 non-matched pairs were also randomly chosen from the Notre Dame dataset.

The image descriptors were feature vectors extracted from image patches as in the method for human detection above. Image patch pairs for which the descriptor vectors were sufficiently close were classified as matched. We computed the Euclidean distance between descriptors of image patch pairs and then, for evaluating performance, an ROC curve was constructed, based on the two histograms of the distances for all true matching and non-matching cases in the dataset. In the learning phase, the parameters of the descriptors were appropriately determined according to the evaluation results of the ROC for the training dataset; minimizing the FP rate when the TP rate is 0.98. After descriptors were learned, performances were evaluated on the test dataset.

The image descriptor was constructed as follows: First, the image patch was smoothed by the Gaussian kernel of the standard deviation $\sigma$, and then the feature was extracted with spatial binning. Finally, L2-Hys of the threshold $\gamma$ was applied to whole feature vector. In the proposed method, the $0^{\text{th}}$order and $1^{\text{st}}$order components were weighted by $\mu$ and $(1-\mu)$, respectively, for calculating distances between descriptors. We applied only GLAC of the orientation bin $D=8$ according to the comparison between GLAC and NLAC in Sec.5.1, and GLAC is compared with the other methods: SIFT [12], Steerable Filter [17], SLAC, and the normalized sum of squared differences (NSSD). The parameters to be learned were $\sigma, \Delta r, \gamma$ in GLAC and SLAC while those of Steerable Filter were $\sigma, \gamma$. The parameters of SIFT and NSSD were set to the values described in [17]. Unlike [17], spatial binning of all methods was not learned but constant $(4\times4)$ in order to accurately compare the performance of the feature extraction methods themselves. Fig. 7(a) shows the results. GLAC outperformed the other methods including SIFT. It is noteworthy that, in all experiments, the learned weight $\mu$ of the $0^{\text{th}}$order was 0 and so the $0^{\text{th}}$order components are redundant for image patch matching as well as for human detection. The learned value of $\sigma$ was non-zero and it is found that preprocessing of smoothing images contributes to an improvement in contrast to human detection. Furthermore, the spatial interval learned was larger ($\Delta r \sim 10$), thus implying a stronger effect for somewhat broader texture alignment. The performance of GLAC




(a) Overall performance (b) Gradient method

(c) Orientation bins (d) Spatial interval

(e) Weighting (f) Auto-correlation order

(g) Normalization (h) Spatial bins

**Fig. 8.** The detection performances of GLAC with various settings. Details are in the text.

along with spatial binning is shown in Fig. 7(b). Finer binning than $2\times2$ produced a superior result to SIFT with $4\times4$ binning.

## 6 Conclusions

We have proposed two methods for extracting image features: Gradient Local Auto-Correlation (GLAC) and Normal Local Auto-Correlation (NLAC). This framework is based on spatial and orientational auto-correlations of local image gradients and normals, which renders shift-invariance and additivity as in HLAC [15]. The gradient is sparsely described in terms of magnitude and orientation for GLAC. The gradients can be extended to normal vectors on the image surface for NLAC. These methods extract local geometrical characteristics of the image surface in more detail than standard histogram based methods, since 2 $^{nd}$order statistics are utilized. In experiments for human detection and image patch matching, the proposed methods produced favorable results compared with the other methods. It was also found that GLAC works well with spatial binning and normalization, although shift-invariance is lost, whereas NLAC without these processings is suitable for shift-invariant recognition problems.

## References

1. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
2. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 878–885 (2005)
3. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence 27, 1615–1630 (2005)
4. Lin, Y.Y., Liu, T.L., Fuh, C.S.: Local ensemble kernel learning for object category recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
5. Smeulders, A.W., Worring, M., Sntini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence 22, 1349–1380 (2000)
6. Boiman, O., Irani, M.: Detecting irregularities in images and in video. In: International Conference on Computer Vision, pp. 462–469 (2005)
7. Belongie, S., Malik, J., Puzicha, J.: Matching shapes. In: International Conference on Computer Vision, pp. 454–461 (2001)
8. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2007)
9. Laptev, I., Lindeberg, T.: Space-time interest points. In: International Conference on Computer Vision, pp. 432–439 (2003)
10. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. International journal of computer vision 73, 213–238 (2007)
11. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: International Conference on Computer Vision, pp. 1–8 (2007)
12. Lowe, D.: Distinctive image features from scale invariant features. International Journal of Compuater Vision 60, 91–110 (2004)

13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 20–25 (2005)
14. Rautkorpi, R., Iivarinen, J.: A novel shape feature for image classification and retrieval. In: International Conference on Image Analysis and Recognition, pp. 753–760 (2004)
15. Otsu, N., Kurita, T.: A new scheme for practical flexible and intelligent vision systems. In: IAPR Workshop on Computer Vision (1988)
16. Russ, J. (ed.): The Image Processing Handbook. CRC Press, Boca Raton (1995)
17. Winder, S., Brown, M.: Learning local image descriptors. In: Computer Vision and Pattern Recognition, pp. 1–8 (2007)
18. Vapnik, V. (ed.): Statistical Learning Theory. Wiley, Chichester (1998)
19. Freeman, W., Adelson, E.: The design and use of steerable filters. Pattern Analysis and Machine Intelligence 13, 891–906 (1991)
20. Mohan, A., Papageorgiou, C., Poggio, T.: Example-based object detection in images by components. Pattern Analysis and Machine Intelligence 23, 349–361 (2001)