Abstract—The Laplacian of Gaussian (LoG) filter is widely used in interest point detection. However, low-contrast image structures, though stable and significant, are often submerged by the high-contrast ones in the response image of the LoG filter, and hence are difficult to be detected. To solve this problem, we derive a generalized LoG filter, and propose a zero-norm LoG filter. The response of the zero-norm LoG filter is proportional to the weighted number of bright/dark pixels in a local region, which makes this filter be invariant to the image contrast. Based on the zero-norm LoG filter, we develop an interest point detector to extract local structures from images. Compared with the contrast dependent detectors, such as the popular scale invariant feature transform detector, the proposed detector is robust to illumination changes and abrupt variations of images. Experiments on benchmark databases demonstrate the superior performance of the proposed zero-norm LoG detector in terms of the repeatability and matching score of the detected points as well as the image recognition rate under different conditions.

Index Terms—Interest point detection, Laplacian of Gaussian filter, zero-norm, image matching, face recognition.

I. INTRODUCTION

Interest point detection is an important research topic in image processing, computer vision and pattern recognition as it provides an effective way to represent images by sparse local patches [1]–[7]. Interest points, such as corners, blobs and stable regions, are the image patterns that differ from their immediate neighborhoods [4]. They are well suited to deal with the challenges of clutter, occlusion and variation of viewing conditions [8]. In the last decades, various detectors have been developed [9]–[26] to extract interest points and widely used in image retrieval [27], image registration [28], panoramic image stitching [29], object categorization [6], object recognition [30], [31], texture classification [32], video shot retrieval [7], build detection [33], [34] and face recognition [14], [35]. However, a key challenge remains as to how to extract stable interest points under illumination changes and abrupt variations that widely exist in natural images.

Existing detectors are usually developed on the basis of local image contrast [9]–[14], such as the Harris [9], [10], Harris-Laplace/Affine [11], Hessian-Laplace/Affine [11], Scale Invariant Feature Transform (SIFT) [12] and Speeded Up Robust Features (SURF) [13] detectors. The Harris and its derivative Harris-Laplace/Affine detectors use the first order derivatives of the image intensity to form the second moment matrix. The square of the trace and the determinant of this matrix are linearly combined to measure the principal intensity changes in two orthogonal directions of an image patch. Therefore, the measurement is proportional to the local contrast, so that low contrast structures will give low responses. As the first order derivative amplifies the image noise, this makes the Harris-Laplace/Affine detectors sensitive to noises in detecting the low contrast structures. The Hessian-Laplace/Affine detectors utilize the second order derivative to create the Hessian matrix for interest points detection. Similar to the Harris detector, the Hessian matrix-based detectors focus on the image structures with large local contrast. Meanwhile, several detectors have been proposed based on the Laplacian of Gaussian (LoG) filter. The SIFT detector employs the difference of Gaussian (DoG) filter to approximate the normalized LoG filter. This method significantly reduces the computational complexity. The SURF detector uses the box filters and the integral images to further accelerate the Hessian-Laplace detector. The Rank Order LoG (ROLG) detector [14] utilizes the rank order filter instead of the linear one to reduce the influence of noises and nearby structures. However, these detectors work well only for structures with high contrast.

Another approach to detect interest points is designed on the basis of image segmentation algorithms or local statistical properties instead of the image contrast. The Maximally Stable Extremal Regions (MSER) [15], [16], Maximally Stable Colour Regions (MSCR) [17], Principal Curvature-Based Region (PCBR) [18] and Boundary Preserving Local Regions (BPLR) [19] detectors extract local structures with watershed-like segmentation algorithms. These detectors face the challenges that image segmentation is, at times, unreliable because of the poor imaging conditions such as image blurring [4]. The statistical properties of local regions are employed by the Smallest Univalence Segment Assimilating Nucleus (SUSAN) [20], Features from Accelerated Segment Test (FAST) [21], and salient region [22] detectors. Both the SUSAN and FAST detectors compute the similarity between the nucleus and its surrounding pixels to generate the

Contrast Invariant Interest Point Detection by Zero-Norm LoG Filter

Zhenwei Miao, Xudong Jiang, Senior Member, IEEE, and Kim-Hui Yap, Senior Member, IEEE
corner maps. The corner maps of both detectors are generated from local regions with fixed size, thus these two detectors are not scale invariant. Another detector focusing on salient region [22] employs the entropy of the intensity histogram of local region to extract interest point. The number of points detected by this detector is small as the greedy cluster method is used to group the nearby interest points [4]. Self-similarity of image patches is computed in [23] to detect the local structures. The similarity which is estimated by the variance of local pixel values can be easily affected by the abrupt changes within the image patch.

Instead of using either image segmentation algorithms (e.g. MSER detector [15]) or local statistical properties (e.g. salient region detector [22]), we propose a Generalized Laplacian of Gaussian (GLoG) filter to tackle the problem caused by illumination changes. The GLoG filter is inspired by the positive properties of the linear LoG filter. Rather than directly using the weighted average of pixel values in a local region as the filter output, the GLoG filter introduces a general location-based gray value estimation function \( \beta(x) \) at location \( x \) to compute the local contrast of each pixel, and a nonlinear factor \( \alpha \) to adjust the influence of the local contrast.

The choice of both the estimation function and the nonlinear factor yields a filter output that is robust to the abrupt changes in a wide range of image variations are evaluated, concerning the scale, viewpoint, blur and rotation changes. As a practical application, we test the detectors in face recognition.

II. THE GLoG FILTER

The GLoG filter is proposed following the analysis of the LoG filter. Properties of the GLoG filter are discussed in this section.

A. Problem Formulation

The LoG function (shown in Fig. 1(a)) is defined as

\[
\omega(x, \sigma) = -\frac{1}{\sigma^2} \left[ 1 - \frac{|x|^2}{2\sigma^2} \right] e^{-\frac{|x|^2}{2\sigma^2}},
\]

where \( x = (x, y) \) is the image spatial coordinate and \( \sigma \) (named as scale factor) is the standard deviation of the corresponding Gaussian function. The coefficient \( \omega(x, \sigma) \) decreases by increasing \(|x|\) for \(|x| > 2\sigma\). Fig. 1(a) reveals that the coefficients of the LoG filter approach zero if their distances from the filter center are large. Therefore, we use a finite sized LoG filter for further analysis. A finite sized LoG filter mask can be divided into two parts according to the sign of the coefficients, as shown in Fig. 1(b). \( S_1 \) is the inner region with the negative sign, and \( S_1 \) corresponds to its surrounding region containing positive weights.

Let \( W_1(\sigma) = \sum_{m \in S_1} \omega(m, \sigma) \) and \( W_2(\sigma) = \sum_{m \in S_2} |\omega(m, \sigma)| \) be the summation of the absolute value of the weights in \( S_1 \) and \( S_2 \), respectively, and \( \mu_1(x) \) and \( \mu_2(x) \) be the weighted mean of the pixel values in \( S_1 \) and \( S_2 \), respectively, defined as

\[
\mu_1(x) = \frac{1}{W_1(\sigma)} \sum_{m \in S_1} \omega(m, \sigma) I(x - m)
\]

and

\[
\mu_2(x) = \frac{1}{W_2(\sigma)} \sum_{m \in S_2} |\omega(m, \sigma)| I(x - m).
\]

Substituting (3) and (4) into (2) yields

\[
r(x, \sigma) = \frac{1}{2} \sum_{m \in S_1} \omega(m, \sigma) (I(x - m) + \mu_1(x))
- \frac{1}{2} \sum_{m \in S_2} |\omega(m, \sigma)| (I(x - m) + \mu_2(x)).
\]
\( w(\mathbf{m}, \sigma) \) for a fixed \( \mathbf{x} \), we have
\[
\sum_{m \in S_1} w(\mathbf{m}, \sigma)\mu_1(\mathbf{x}) = \sum_{m \in S_2} |w(\mathbf{m}, \sigma)|\mu_1(\mathbf{x}), \tag{6}
\]
and
\[
\sum_{m \in S_2} |w(\mathbf{m}, \sigma)|\mu_2(\mathbf{x}) = \sum_{m \in S_1} w(\mathbf{m}, \sigma)\mu_2(\mathbf{x}). \tag{7}
\]
Substituting (6) and (7) into (5) and simplifying, it gives
\[
r(\mathbf{x}, \sigma) = \frac{1}{2} \sum_{m \in S_1} w(\mathbf{m}, \sigma)(I(\mathbf{x} - \mathbf{m}) - \mu_1(\mathbf{x})) + \frac{1}{2} \sum_{m \in S_2} |w(\mathbf{m}, \sigma)|(\mu_1(\mathbf{x}) - I(\mathbf{x} - \mathbf{m})). \tag{8}
\]
\( \mu_1(\mathbf{x}) \) and \( \mu_2(\mathbf{x}) \) can be interpreted as the location-based gray value estimates from the pixels in \( S_1 \) and \( S_2 \), respectively. With this interpretation, \( I(\mathbf{x} - \mathbf{m}) - \mu_i(\mathbf{x}), i \in \{1, 2\} \), is the contrast between the pixel at \( \mathbf{x} - \mathbf{m} \) and the location-based gray value estimate of its counterpart region. The first term in (8) is the weighted contrast between the pixels in \( S_1 \) and the estimate of \( S_2 \), and the second term is the weighted contrast between the estimate of \( S_1 \) and the pixels in \( S_2 \). The LoG filter output is the average of these two local contrasts.

In (8) we observe the following two properties of the linear LoG filter-based detectors.

1) The response of the LoG filter is proportional to the local intensity contrast. Regions with high contrast receive high responses. For example, Fig. 2(a) contains four Gaussian blobs of the same scale but different amplitudes of 255, 128, 64 and 32, respectively. The pixel values are quantized to the integer in the range of \([0 \ 255]\) as that of the real digital image. All of the filter responses in this paper are normalized by their peaks for a better comparison. Fig. 2(b) shows the profile of the LoG response across the center of each blob, of which the responses are proportional to the amplitude of the blobs. The LoG filter does not amplify the response of the low contrast image structures, making the LoG filter-based detectors difficult to detect these structures.

2) The LoG filter-based detectors perform better in the simple round blobs rather than irregular image structures or the ones in a cluttered background. For example, Fig. 3 shows results of the SIFT detector. The disk is successfully detected in Fig. 3(a) while it fails in Fig. 3(b). The reason can be explained by (8). Eq. (8) shows that the LoG filter uses the weighted mean of the pixel intensities, \( \mu_1 \) and \( \mu_2 \), as the location-based gray value estimates. It is well known that the mean is optimal for Gaussian distributed data. However, for most natural images, Gaussian noise is much weaker compared to the image structures. The irregular image structures have much higher influence on the intensity estimates of \( S_1 \) and \( S_2 \). In this case, the weighted mean adversely affect the performance of the detector. Fig. 3(c) and 3(d) further illustrate our analyses. The existence of the smaller darker structure significantly affects the gray value estimate of the disk’s surrounding region. This leads to significant deviation of the DoG responses in detecting the larger scale disk, which are plotted in Fig. 3(c) and Fig. 3(d). As a result, the SIFT detector fails to detect the disk in Fig. 3(b) though it can successfully detect that in Fig. 3(a).

The above two properties show that the LoG filter has two problems in detecting blobs: 1) difficult in detecting low contrast structures because its response depends heavily on the intensity contrast, and 2) sensitive to the nearby structures given that it uses the weighted mean as the location-based gray value estimate in the corresponding parts. Solving these two problems is critical for robust local structure detection. Therefore, we propose an alternative filter which is elaborated in the following section.

B. The Proposed GLoG Filter

Instead of directly using the weighted mean of the local contrast as the filter output (8), we propose to apply nonlinear operations to the local intensity contrast. Specifically, the GLoG filter is generalized from (8) as below
\[
\begin{align*}
  r_G(\mathbf{x}, \alpha, \beta_1(\mathbf{x}), \beta_2(\mathbf{x})) \\
  = \sum_{\mathbf{m} \in S_1} w(\mathbf{m})\text{sgn}(I(\mathbf{x} - \mathbf{m}) - \beta_1(\mathbf{x}))|I(\mathbf{x} - \mathbf{m}) - \beta_2(\mathbf{x})|^\alpha \\
  + \sum_{\mathbf{m} \in S_2} |w(\mathbf{m})|\text{sgn}(\beta_1(\mathbf{x}) - I(\mathbf{x} - \mathbf{m}))|\beta_1(\mathbf{x}) - I(\mathbf{x} - \mathbf{m})|^\alpha.
\end{align*} \tag{9}
\]

For notation simplicity, the scale $\sigma$ and constant 1/2 are omitted without loss of generality. $\beta_1(x)$ and $\beta_2(x)$ are the corresponding location-based gray value estimates from the pixels in the outer and inner regions, and $\text{sgn}(x)$ is the signum function defined as

$$\text{sgn}(x) \triangleq \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0. \end{cases} \quad (10)$$

The GLoG filter differs from the linear LoG filter mainly in two aspects. Firstly, the weighted mean ($\mu_i(x)$, $i \in \{1, 2\}$) is replaced with a more general location-based gray value estimation function, ($\beta_i(x)$, $i \in \{1, 2\}$). This general estimation function can be selected to be the weighted mean, the weighted median or the weighted iterative truncated/trimmed mean [37]–[40] that owns some merits of both the weighted mean and median. Secondly, the difference between the pixel value and the corresponding location-based gray value estimate is replaced with the absolute value of the difference raised to the power of $\alpha$, and it is then multiplied with the sign of the difference. By uncoupling the sign from the contrast amplitude, $\alpha$ can be assigned to any value as a means to alter the filter characteristics. In the following, analyses of the GLoG filter are given.

For the case $\alpha = 1$, the GLoG filter (9) degrades to

$$r_G(x, 1, \beta_1(x), \beta_2(x)) = \sum_{m \in S_1} w(m)(I(x - m) - \beta_2(x)) + \sum_{m \in S_2} |w(m)|(|\beta_1(x) - I(x - m)|) \quad (11)$$

which is the linear combination of the inner and surrounding local contrasts. From (11), Property 1 is straightforward.

Property 1: The GLoG filter with $\alpha = 1$ turns into the linear LoG filter if $\beta_1(x)$ and $\beta_2(x)$ are the weighted means $\mu_1(x)$ and $\mu_2(x)$, respectively, defined as

$$r_G(x, 1, \mu_1(x), \mu_2(x)) = \sum_{m \in S_1} w(m)(I(x - m) - \mu_2(x)) + \sum_{m \in S_2} |w(m)|(|\mu_1(x) - I(x - m)|) \quad (12)$$

$$= 2W_1(\sigma)(\mu_1(x) - \mu_2(x)) \quad (13)$$

which is achieved by substituting (3) and (4) into (12).

Next we analyze the property of the GLoG filter against the scale factor $\sigma$. It can be proved that

$$W_1(\sigma) = 2 \exp(-1)/\sigma^2. \quad (14)$$

As $W_1(\sigma)$ is negatively correlated with $\sigma^2$, it is likely that the response of the LoG filter decreases as the scale increases. Given that the set of GLoG filters are built on the LoG filter, these filters inherit this property. The responses of the mean-based GLoG filters with respect to the scale are shown in Fig. 4. It demonstrates the decreasing values of the filter responses against the increasing scale. Therefore, normalization is needed for the scale selection [41]. The normalized GLoG filter is defined as

$$r_{nG}(x, \alpha, \sigma, \beta_1(x), \beta_2(x)) = r_G(x, \alpha, \sigma, \beta_1(x), \beta_2(x))/W_1(\sigma). \quad (15)$$

The responses are shown in Fig. 5. It shows that the GLoG filter responses reach the global peak when the scale of the GLoG filter matches with that of the blob structures.

In addition to the weighted mean, an alternative choice of the location-based gray value estimate is the weighted median [39] that is robust to abrupt variations and impulsive noise. The weighted medians of the pixel intensities in regions $S_1$ and $S_2$ are defined as

$$\phi_1(x) \triangleq \text{median}\{|w(m) \odot I(x - m)|; m \in S_1\} \quad (16)$$

and

$$\phi_2(x) \triangleq \text{median}\{|w(m) \odot I(x - m)|; m \in S_2\}, \quad (17)$$

where $\odot$ is the replication operator defined as

$$w_1 \odot x \triangleq x, x, \ldots, x. \quad (18)$$

Property 2: The output of the GLoG filter with $\alpha = 1$ and $\beta_1(x) = \phi_1(x)$ and $\beta_2(x) = \phi_2(x)$ is a linear combination of
The weighted mean and weighted median, as

\[ r_G(x, 1, \phi_1(x), \phi_2(x)) = \sum_{m \in S_1} w(m)(I(x - m) - \phi_2(x)) + \sum_{m \in S_2} |w(m)|(\phi_1(x) - I(x - m)) = W_1(\mu_1(x) - \mu_2(x) + \phi_1(x) - \phi_2(x)). \]  

(19)

In contrast to the linear LoG filter, the GLoG filter with the median compromises between the linear and median filters. Hence, it improves the filter’s performance in noise suppression and structure preservation.

**Property 3:** The nonlinear factor \( \alpha \) controls the influence of the local contrast to the filter output. The response of the GLoG filter is less sensitive to the contrast for a smaller \( \alpha \) value.

Setting \( \alpha = 1 \), the responses of the mean- and median-based GLoG filters are proportional to the local image contrast. This property is shown in Fig. 2(b) and Fig. 7(a). Decreasing \( \alpha \) enhances the response of the low contrast structures. Examples for the mean-based GLoG filters are shown in Fig. 6(a), (b) and (c) with \( \alpha \) equal to 0.5, 0.25 and 0.125, respectively. A similar trend appears in the responses of the median-based GLoG filters shown in Fig. 7(b), (c) and (d). Compared to the mean-based GLoG filters, the median-based GLoG filters produce much weaker responses surrounding blobs, which will reduce the spurious peaks being detected.

These observations inspire us to design the \( \ell_0 \)-LoG detector. We will elaborate it in the following section.

### III. The \( \ell_0 \)-LoG Detector

The \( \ell_0 \)-LoG filter is introduced and its properties are analyzed. Following that we propose the \( \ell_0 \)-LoG detector.

#### A. The Proposed \( \ell_0 \)-LoG Filter

Section II-B reveals the function of the nonlinear factor \( \alpha \) in controlling the influence of the intensity contrast to the GLoG filter output. A small \( \alpha \) value enhances the response of the low contrast structures. For any value \( x \), \(|x|^0 = 1 \) if \( x \neq 0 \). Therefore, the output of the GLoG filter is independent of the local image contrast if \( \alpha = 0 \). Inspired by this, the proposed \( \ell_0 \)-LoG filter is defined as

\[
r_{\ell_0}(x, \sigma) \triangleq \sum_{m \in S_1} w_{\ell_0}(m, \sigma) \text{sgn}(I(x - m) - \beta_2(x)) + \sum_{m \in S_2} |w_{\ell_0}(m, \sigma)| \text{sgn}(\beta_1(x) - I(x - m)),
\]

(20)

where \( w_{\ell_0}(m, \sigma) = w(m, \sigma) / W_1(\sigma) \) is the normalized weighting coefficient. Eq. (20) is directly derived from (9) and (15) by setting \( \alpha = 0 \). The output of the \( \ell_0 \)-LoG filter depends only on the weighted number of pixels that are brighter or darker than the intensity estimates of their counterpart regions. This results in the benefit that it is invariant to the image contrast. The responses of the \( \ell_0 \)-LoG filters on the image given in Fig. 2(a) are shown in Fig. 8. It is shown that the filter responses are the same for the four blobs. This verifies the contrast invariant property of the \( \ell_0 \)-LoG filter.

Moreover, Fig. 8 shows that the absolute values of the negative peaks surrounding the blob generated by the median-based \( \ell_0 \)-LoG filter are much lower than those of the mean-based \( \ell_0 \)-LoG filter. As lower peak surrounding the blob leads to a smaller risk of detecting spurious interest points, the median-based \( \ell_0 \)-LoG filter is more robust than the mean-based \( \ell_0 \)-LoG filter. It is well known that the median is much more robust to outlier than the mean [37], [38], [40]. Henceforth, compared to the mean, the difference between a pixel and the median of its counterpart region is more robust to the impulsive noise in that region. As a result, using the median-based \( \ell_0 \)-LoG filter, the detection of a structure will be less affected by the nearby abrupt changes of image intensity. Furthermore, the weighted median estimate is one of the pixel values in the filter window, which enhances noise robustness of the zero-norm LoG filters.
Property 4: For the $\ell_0$-LoG filter, we have $-2 \leq r_{G_0}(x, \sigma) \leq 2$. For a dark region, if all pixels in $S_2$ are darker than the gray value estimate of $S_1$ and all pixels in $S_1$ are brighter than the gray value estimate of $S_2$, i.e.

$$I(x - m) < \beta_1(x), \quad \forall m \in S_2,$$

and

$$I(x - m) > \beta_2(x), \quad \forall m \in S_1,$$

the response of the $\ell_0$-LoG filter reaches the maximum of 2. If the majority of the expanded pixels in $S_2$ (expanded by the corresponding weights) is darker than the gray value estimate in $S_1$ and the majority of the expanded pixels in $S_1$ is brighter than the gray value estimate in $S_2$, $r_{G_0}(x) > 0$. Likewise, the $\ell_0$-LoG filter has the analogous property in bright regions.

B. The Proposed $\ell_0$-LoG Detector

As shown above, the median-based $\ell_0$-LoG filter is invariant to the image contrast and robust to the abrupt changes of the nearby structures. Given such advantage, we design the median $\ell_0$-LoG filter-based detector ($\ell_0$-LoG detector). In order to handle the non-smoothness of the filter response caused by the finite number of pixels in a filter window, a Gaussian filter $G(x, \sigma_0)$ is used

$$\hat{r}_{G_0}(x, \sigma) = r_{G_0}(x, \sigma) \otimes G(x, \sigma_0),$$

(21)

where $\sigma_0$ is set to 1 in this work. By detecting the local peaks on the response image $\hat{r}_{G_0}(x, \sigma)$, interest points with the scale $\sigma$ are extracted. The following two steps remove the unstable points from the ridge and extend the $\ell_0$-LoG detector into multiple scales.

1) Ridge Suppression: The $\ell_0$-LoG filter on ridge is strong if the scale of the $\ell_0$-LoG filter is close to the width of the ridge. Slight vibration of the pixel gray values may cause false detection of interest points on the ridge. Such unstable points are expected to be removed.

Although the peaks of the $\ell_0$-LoG filter output on ridge show large amplitude, the absolute value of the peak $|\hat{r}_{G_0}(x)|$ is close to the maximum absolute value in the corresponding surrounding region $S_1$. Hence, we employ the ratio

$$\lambda = \max\{|\hat{r}_{G_0}(x - m, \sigma)|, m \in S_1\} / |r_{G_0}(x, \sigma)|$$

(22)

to remove the unstable interest points on the ridge. If $\lambda$ is close to 1, the peak will be much similar to its nearby region. Such interest point candidate is most likely on the ridge. We remove such candidates if $\lambda < 0.95$, which is chosen experimentally.

2) The $\ell_0$-LoG Detector in Multiple Scales: Interest point detection in multiple scales is an important issue in vision applications. It is known that the local structures exist over a range of scale [22]. In order to detect the scale of the local structures from this range, Lindeberg [41] proposed a set of matching filters that later has been adopted by several detectors such as SIFT [12], SURF [13] and Harris/Hessian-Laplace [11]. The filter response is maximized once the scale of the filter matches with the scale of a local structure. Therefore, the detected scale of the local structure is set to be proportional to the scale of the matched filter. The matching filters chosen in [41] consist of a set of normalized LoG filters with a varying scale $\sigma$. These filters are optimal in detecting the structures of Gaussian shape. For structures of other shapes, the peak may become flat or even multiple peaks. In practice, the shapes of local structures appear in different ways, and most of them are different from Gaussian shape. It is often difficult to get a stable local maximum in both scale and spatial dimensions. As such, using the local maximum along the scale dimension to determine the scale is unreliable. The problem of scale selection also happens when using the MSER detector [15]. This detector employs the local minima of the area-changing rate of the extremal regions along the intensity levels to determine the scale. Sometimes, no sharp valley exists along the scale dimension, so that a slight vibration may influence the scale selection. For example, Fig. 9 shows the scales detected by the MSER detector on a Gaussian blob. Multiple scales are detected on the single blob.

By changing the scale $\sigma$ of the filter mask, the $\ell_0$-LoG detector achieves the multi-scale detection of local structures. Similar to other detectors, the structures detected by the $\ell_0$-LoG detector appear in a wide range of scales. Fig. 10 shows the normalized LoG and $\ell_0$-LoG filter responses at the center of a dark Gaussian blob along the scale dimension. It is shown that the $\ell_0$-LoG filter response is flatter than the normalized LoG filter response. No sharp maximum is generated. Therefore, the local extremum along the scale dimension is sensitive to noise, and unreliable in determining the scale size. From another perspective, as the local structure appears in multi-scales, the result should be more reliable if
we group them together instead of selecting a single one to represent the structure. In view of this, a grouping method is proposed as follows to cluster the interest points that come from the same structure.

Interest points from the same structure have the properties that they are all bright or all dark, and are neighbors in both the spatial and scale dimensions. It suggests that these interest points are inter-correlated. Therefore, we name them as connected interest points and define them as follows. Let an interest point at location \((x_i, y_i)\) and scale \(s_i\) be represented by \(P_i = (x_i, y_i, s_i, \tilde{r}_{G0p}, F_i)\) where \(\tilde{r}_{G0p}\) is the \(\ell_0\)-LoG filter response and \(F_i \in \{\text{Bright}, \text{Dark}\}\) is the flag of negative or positive peak. Each \(P_i\) is a local peak in the spatial dimensions. A connection of two interest points \(P_i\) and \(P_j\) is found if:

1) They are both bright (or both dark) regions;
2) They are close in both the spatial and scale dimensions.

In this work, the closeness of two interest points in the spatial dimensions is measured by the ratio between the Euclidean distance of the peak locations and the corresponding scale, defined as

\[
d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} / s_i. \tag{23}
\]

For a connected pair \(P_i\) and \(P_j\), the threshold \(d_{ij}\) should be smaller than 0.3, which is chosen experimentally. In the scale dimension, the scales \(s_i\) and \(s_j\) of a connected pair should either be the immediate or the second nearest neighbor to each other. The group center of the connected interest points is defined as follows. Assume a group \(G_p\) contains \(N\) interest points as \(G_p = \{P_{p1}, P_{p2}, \ldots, P_{pN}\}\). The interest point center \(P_p = (x_p, y_p, s_p, \tilde{r}_{G0p}, F_p)\) for this group is computed by setting the location \(x_p = \text{mean}(\{x_i\}_{i=1}^N)\), \(y_p = \text{mean}(\{y_i\}_{i=1}^N)\), the scale

\[
s_p = \frac{\sum_{i=1}^N \tilde{r}_{G0p1}^2 s_i p_i / \sum_{i=1}^N \tilde{r}_{G0p1}^2}{\sum_{i=1}^N \tilde{r}_{G0p1}^2}, \tag{24}
\]

and the response

\[
\tilde{r}_{G0p} = \max(\tilde{r}_{G0p1}, \tilde{r}_{G0p2}, \ldots, \tilde{r}_{G0pN}). \tag{25}
\]

By weighting the scale with \(\tilde{r}_{G0p1}^2\), the scales with large \(\ell_0\)-LoG filter responses have high influence on determining the scale of the corresponding group.

The proposed algorithm for the \(\ell_0\)-LoG detector is summarized as follows:

1) Compute the median-based \(\ell_0\)-LoG filter responses on multi-scales with \(20\), and smooth the filter response with \(21\).
2) Detect the local extremes in the spatial space for each scale.
3) Remove the points on ridges with \(22\).
4) Group the interest points that correspond to the same structure with the above proposed criterion.

IV. EXPERIMENTS

In this section, we evaluate the \(\ell_0\)-LoG detector, and compare it with the start-of-the-art detectors. A visual comparison is given in Section IV-A to illustrate the performance of the \(\ell_0\)-LoG detector under the illumination changes.

Experiments (Section IV-B to Section IV-C) are designed to test the repeatability and the discrimination of the detected interest points. For simplicity, HR-A and HS-A stand for the Harris-Affine detector [11] and Hessian-Affine detector [11], respectively. To evaluate the discriminative ability, the SIFT descriptor [12] is applied to all detected interest points.

A. Visual Inspection

Images from the Oxford database [3] ‘leuven’ data set are used to show examples of visual results of the \(\ell_0\)-LoG detector under illumination changes. The parameters are adjusted to show the case where the same numbers of points are roughly detected from the image Fig. 11(a) by all detectors. Results are shown in Fig. 11(b), (c) and (d). Then, the detectors with the same parameters are applied to the image Fig. 11(e) that has a different illumination of Fig. 11(a) in the Oxford database. Results are shown in Fig. 11(f), (g) and (h). It is observed that the numbers of the interest points detected by the SIFT detector (shown in Fig. 11(b) and (f)) and the MSER detector (shown in Fig. 11(c) and (g)) decrease when the scene darkens. A number of interest points that are detected in the first image are not detected in the second one. In contrast, the illumination change has little effect in the \(\ell_0\)-LoG detector. Most of the structures detected by the \(\ell_0\)-LoG detector are repeated in both illumination settings (shown in Fig. 11(d) and (h)).

B. Test on Oxford Database

The aim of this experiment is to evaluate the detectors under different variations based on the protocol in [3]. For a stable interest point detector, the detected points should have a high repeatability which reveals the stability of the detector under different image variations. In addition, a high matching score is desirable as it indicates a more powerful discrimination. Therefore, we use the repeatability and the matching score to evaluate the detectors. Two detected regions are repeated if their overlap is above a certain threshold (it is set to be 60\% of the union of their detected regions as suggested in [3]). Two detected regions are matched if 1) they are repeated, and 2) their descriptors are the nearest-neighbor in the descriptor space. The repeatability is defined as the ratio between the number of repeated points and the larger number of detected points in the common scene and scales of the image pair. The matching score is defined as the ratio between the number of matched points and the larger number of detected points in the common scene and scales of the image pair.

The test data sets are from the standard and publicly available Oxford database [3]. Eight data sets are contained in this database, namely ‘boat’, ‘bark’, ‘graf’, ‘wall’, ‘ubc’, ‘bikes’, ‘trees’ and ‘leuven’. These data sets include five types of image variations: scale changes (‘boat’ and ‘bark’), viewpoint variation (‘graf’ and ‘wall’), JPEG compression (‘ubc’), image blur (‘bikes’ and ‘trees’) and lighting variation (‘leuven’). Each data set contains 6 images with 5 different levels of variations between the first image and the other five images.

Similar to what is done in [23], interest points are detected in the half-sampled images. For the \(\ell_0\)-LoG detector,
Fig. 11. (a) and (e): two input images with different illuminations. (b) and (f): interest points detected by the SIFT detector. (c) and (g): interest points detected by the MSER detector. (d) and (h): interest points detected by the proposed $\ell_0$-LoG detector. The repeated interested points are plotted in yellow color.

| TABLE I |
| Number of Detected Points on the First Image of Each Data Set |
|---------|---------|---------|---------|---------|---------|
| $\ell_0$-LoG | SIFT | HR-A | HS-A | MSER |
| boat     | 1584   | 1501  | 1549  | 1429   | 1324   |
| bark     | 1386   | 1507  | 1209  | 1513   | 1681   |
| graf     | 1468   | 1347  | 1480  | 1495   | 1396   |
| wall     | 1535   | 1460  | 1520  | 1568   | 1593   |
| ubc      | 1604   | 1495  | 1460  | 1330   | 1565   |
| bikes    | 1545   | 1456  | 1515  | 1499   | 1401   |
| trees    | 1443   | 1557  | 1480  | 1515   | 1487   |
| leuven   | 1531   | 1426  | 1476  | 1501   | 1648   |

interest points are detected in 5 octaves by half-sampling the previous octave. In each octave, local extrema are detected in 3 scales: $\{\sigma_n\}_{n=1,2,3} = \{1.6 \times 2^{1/3}, 1.6 \times 2^{2/3}, 1.6 \times 2\}$. Four detectors, the MSER [15], HR-A [11], HS-A [11] and SIFT [12] detectors are compared with the $\ell_0$-LoG detector. The codes of the HR-A, HS-A and MSER detectors are obtained from http://www.robots.ox.ac.uk/~vgg/research/affine/evaluation.html. The code of the SIFT detector is obtained from http://www.robots.ox.ac.uk/~vedaldi/assets/sift/versions/. For each data set, the detector parameters are adjusted so that all detectors output roughly the same number of interest points on the first image as shown in Table I. Parameters which are tuned include the response threshold of the $\ell_0$-LoG detector, the cornerness thresholds of the HR-A and HS-A detectors, the strength threshold of the SIFT detector, and the minimum margin of the MSER detector. Parameters of each detector tuned in the first image are applied to the rest images in each data set.

The repeatability and matching score are shown in Fig. 12. The 8 columns correspond to the 8 data sets. In each column, the horizontal axis represents the image index with increasing variation to the first image.

In respect of scale change, the 1st and 2nd columns of Fig. 12 show that the $\ell_0$-LoG detector performs the best for both the structure scene (‘boat’) and the textured scene (‘bark’).

Regarding viewpoint change, the 4th column of Fig. 12, on the textured scene (‘wall’) shows that the $\ell_0$-LoG detector yields the best performance. For the structure scene (‘graf’), however, the MSER detector outperforms the $\ell_0$-LoG detector for large viewpoint change. This is attributed to the affine invariant property of the MSER detector.

For JPEG compression (‘ubc’), the HS-A and HR-A detectors show the highest repeatability and matching score. JPEG compression generates many artificial low contrast structures which are detected by the $\ell_0$-LoG detector. This causes lower repeatability and matching score of the $\ell_0$-LoG detector compared to the HS-A and HR-A detectors.

The results for images blurring are shown in the 6th and 7th columns of Fig. 12. As image blurring changes the contrast of different structures differently, the $\ell_0$-LoG detector outperforms others on both the structured scene (‘bikes’) and the textured scene (‘trees’).

The 8th column of Fig. 12 shows the results of the illumination change (‘leuven’). As the $\ell_0$-LoG filter is independent to the local image contrast, it is not surprising that the proposed detector significantly outperforms the other detectors under the illumination variation.

We use the ‘leuven’ data set to evaluate the localization error of the interest points. The average localization errors of the $\ell_0$-LoG, SIFT, HR-A, HS-A and MSER detectors...
Fig. 12. (a) Repeatability and (b) matching score of detected points on the Oxford database. In each column, the horizontal axis represents the image index in the corresponding data set.

Fig. 13. Repeatability of the weighted median and weighted mean based $\ell_0$-LoG detectors on the 'leuven' data set.

are 1.7, 1.4, 2.5, 2.3 and 1.3, respectively. It is shown that the localization error for the $\ell_0$-LoG is smaller than that of the HR-A and HS-A detectors but larger than that of the SIFT and MSER detectors. In practice, such a small difference of localization accuracy makes minimal effect in image matching.

Experiments are also carried out to evaluate the weighted mean-based $\ell_0$-LoG detector and the weighted median-based $\ell_0$-LoG detector on the ‘leuven’ data set. The repeatability is shown in Fig. 13. The weighted median-based $\ell_0$-LoG detector consistently achieves a higher repeatability than that of the weighted mean-based detector. The reason is that, to detect a point from an image, the nearby image structures that may adversely affect the detection are most likely to be impulse-based. Therefore, in general the weighted median estimator is a better choice.

C. Application to Face Recognition

Face recognition is an active research topic [42]. Although most approaches are based on the dimension reduction [43], [44] or sparse representation [45], interest points are recently utilized in some techniques for face recognition [46], [47]. We evaluate the proposed $\ell_0$-LoG detector on face databases for three reasons. First, many face databases are publicly available, providing a convenient and rich test set for evaluation. Second, face images are featured by both high and low contrast regions, which is one of the best applications to evaluate the effectiveness of detectors. Third, the ground-truth of the databases for face identification is unambiguous. It is crystal clear whether an identification result is correct or wrong. As all detectors with the default setting supplied by the authors produce too few interest points in face images to identify face reliably, the contrast threshold that is used to remove the low contrast interest points is set to be zero for all detectors. This greatly improves the face recognition accuracy of all detectors. For the MSER detector, the minimum size of output region is decreased from 30 to 5 and the maximum relative region size is increased from 0.01 to 0.1 to make it detect roughly the same number of interest points as that of the
the variations of the test images are well represented by the gallery images. Consequently, the $\ell_0$-LoG, SIFT and MSER detectors achieve high recognition rates. The HR-A detector gives the worst performance because human face is a non-rigid surface, and there are few sharp corners in a face image. Although MSER detects more than double the number of points by SIFT, its recognition rate is lower than that of SIFT.

2) Test on ORL Database: Images in the ORL [42] database are normalized into the size of $50 \times 57$. The first 5 images of all 40 subjects are chosen as the gallery set, and the remaining 5 images as the probe set. The repeatability, number of matched points and the first ranked image recognition rate are given in Table III. The cumulative image matching curve is shown in Fig. 14(b). Although the ORL database has fewer subjects than the AR database, the performances of all the detectors here except MSER are worse as compared to those on the AR database. This could be accounted for by the fact that the smaller image size in the ORL database on which the information captured by the detectors is less than that on the AR database. Nevertheless, Table III and Fig. 14(b) show that the performance of the $\ell_0$-LoG detector is much better than those of other detectors. Although the MSER detector captures the most number of interest points on the ORL, GT, and FERET databases, its recognition rate is lower than that of the $\ell_0$-LoG detector as shown in Table III-V and Fig. 14(b)-14(d).

3) Test on GT Database: Images in the GT database [42] have large variations regarding expression, pose and illuminations. Gray images in this database are normalized into the size of $60 \times 80$. The first 8 images of all 50 subjects are chosen as the gallery set, and the remaining 7 images as the probe set.

The experimental results are given in Table IV and Fig. 14(c). The cumulative matching curve of the HR-A detector is not shown in Fig. 14(c) and Fig. 14(d) because it is drastically lower than that of the other detectors in these figures. Due to the large variations of the images in the GT database, the performances of all the detectors

$\ell_0$-LoG detector. All the detected interest points are described by the SIFT descriptor. The algorithm of image recognition through interest point matching given in [12] is applied.

1) Test on AR Database: Gray images in the AR [48] database are normalized into the size of $60 \times 85$. In total, 75 subjects with 14 nonoccluded images per subject are used in the experiment. The first 7 images of all subjects are chosen as the gallery set, and the remaining 7 images as the probe set.

In addition to the repeatability used in Section IV-B, the number of matched points, the first ranked image recognition rate and the cumulative image matching curves, which show the top $k$ recognition rate, are used to evaluate the detectors. The experimental results are shown in Fig. 14(a) and Table II. Table II reveals that the $\ell_0$-LoG detector achieves a much higher repeatability than other detectors. The number of matched points for the $\ell_0$-LoG detector is much larger than other detectors. Therefore, it is not a surprise that the image recognition rate of the $\ell_0$-LoG detector is the best among the five detectors. As images in the AR database are taken under controlled conditions of the illumination and viewpoints [48],
on the GT database are worse compared with those on the AR and ORL databases. However, the $\ell_0$-LoG detector still outperforms the other detectors. Its repeatability and number of matched points are significantly higher than the other detectors. The first ranked image recognition rate of the $\ell_0$-LoG detector is still higher than 90%.

4) Test on FERET Database: Images in the FERET database [49] are cropped into the size of 60 × 80. 1194 subjects with 2 images per subject are selected. The first image of all subjects is chosen as the gallery set, and the remaining images as the probe set.

Experimental result given in Table V shows that the $\ell_0$-LoG detector achieves the higher repeatability and larger number of matched points than other detectors. For such a high quality database, the $\ell_0$-LoG detector significantly outperforms the other detectors over all ranks, as shown in Fig. 14(d).

We evaluate the efficiency of the proposed detector on detecting interest points from face images under the Windows 7 system with the Intel Core i5 CPU 3.2 GHz and 32 GB RAM. The $\ell_0$-LoG detector is implemented in Matlab. The average running time of the $\ell_0$-LoG detector on the AR, ORL, GT and FERET databases is 0.18s, 0.09s, 0.18s and 0.18s, respectively. The SIFT code downloaded from http://www.robots.ox.ac.uk/~vedaldi/assets/sift/versions/ achieves average running time of 0.04s, 0.02s, 0.04s and 0.04s on the AR, ORL, GT and FERET databases, respectively. The $\ell_0$-LoG detector with its current version of implementation without code optimization is about 4.5 times slower than SIFT, which has very efficient implementation with the DoG approximation to the LoG filter.

V. CONCLUSIONS

In this paper, a generalized Laplacian of Gaussian (GLoG) filter is derived from the linear LoG filter, and the $\ell_0$-LoG filter is proposed from the derived GLoG filter. Based on the proposed filter, the $\ell_0$-LoG detector is designed to detect interest points from images. By using the weighted median as the location estimate of the gray values in inner/surrounding regions and employing the weighted number of brighter/darker pixels as the blob map, the $\ell_0$-LoG detector is invariant to the contrast of the local image structures and more robust to the impulsive noise and nearby abrupt structures. Grouping algorithm is developed to select local structures in scale space. Compared to the traditional LoG filter, the $\ell_0$-LoG response is immune to image contrast and robust to cluttered surrounding. Experiments on various data sets demonstrate that the $\ell_0$-LoG detector has better performance in dealing with blurring and illumination changes compared to other four detectors in terms of the repeatability and matching score of the detected points and the image recognition rate.

Similar to other detectors that are designed for blob detection, the proposed filter responses for the blobs are higher than the corners. Nevertheless, the response of the proposed filter also forms a peak or extrema for a corner in the spatial space. As such, the proposed detector is able to detect corners. However, compared to the blob structures, the scale of the corner is difficult to measure as the peak response of the corner appears at slightly shifted spatial positions in different scales. The proposed method can alleviate this problem by grouping multiple spatially close peaks in different scales into one.

REFERENCES
