

The Scale of Edges

Xian-Ming Liu^{1*}, Changhu Wang², Hongxun Yao¹, Lei Zhang²

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

²Microsoft Research Asia, Beijing, China

¹{xmliu, yhx}@hit.edu.cn, ²{chw, leizhang}@microsoft.com

Abstract

Although the scale of isotropic visual elements such as blobs and interest points, e.g. SIFT[12], has been well studied and adopted in various applications, how to determine the scale of anisotropic elements such as edges is still an open problem. In this paper, we study the scale of edges, and try to answer two questions: 1) what is the scale of edges, and 2) how to calculate it. From the points of human cognition and physical interpretation, we illustrate the existence of the scale of edges and provide a quantitative definition. Then, an automatic edge scale selection approach is proposed. Finally, a cognitive experiment is conducted to validate the rationality of the detected scales. Moreover, the importance of identifying the scale of edges is also shown in applications such as boundary detection and hierarchical edge parsing.

1. Introduction

The concept of *scale* in computer vision was first formally discussed in the *scale space* in the 1980s, which was developed by Witkin [19] and Koenderink [7] to describe the image representations at different resolutions. Thus, the scale in the scale space is regarded as a parameter of image resolution [7], based on the fact that an image will derive different levels of details at different resolutions, as shown in Fig. 1.

The scale of isotropic visual elements, such as blobs and interest points, was then studied based on the *isotropic assumption* [19] in the scale-space theory. Lindeberg further analyzed the magnitude changes of local extreme points in the scale space and proposed a normalized scale selection strategy for different detectors [8, 9, 11]. In this way, the scale was further explained as the size of a blob or the diffusion region around an interest point. Afterwards, based on Lindeberg’s formulation, Lowe replaced the Hessian detector in [11] with a Difference of Gaussian (DoG) for efficiency and developed the SIFT feature [12], which is now widely used in computer vision researches.

Although it has been well acknowledged since the suc-

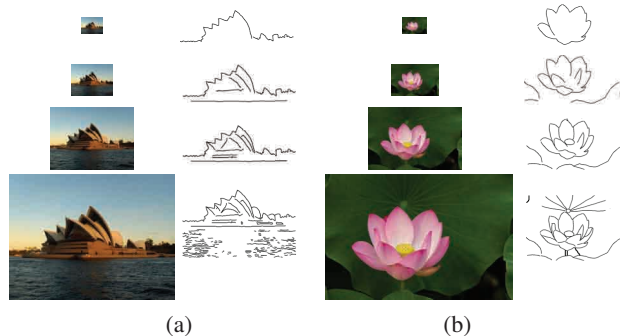


Figure 1. Images at different resolutions will exhibit different levels of edge details. Two examples in the cognitive experiment are shown here. For each image, the subjects were asked to label the most salient edges at each resolution.

cess of SIFT that studying the scale of visual elements is one of the most fundamental prerequisites for feature detection and representation, there is still not a commonly acceptable explanation to the scale of anisotropic elements, such as edges.

Therefore, in this work, we study the scale of edges, and try to answer the following two questions:

1) What is the scale of edges?

Most of existing work related to the scale of edges was studied based on the isotropic assumption [7]. As a natural extension, Lindeberg proposed a scale selection method for edges based on the normalized isotropic scale space [10]. However, as admitted in [10] and also shown in our study, this “scale” is just a *measurement of the diffusion degree of the step edge*, which can be viewed as the “shadow” region of the step edge. Another classical work was proposed by Jurie and Schmid [6], in which the scale of a contour was defined as *the size of a circle which is the most coincident with the contour*, but it only works for contours near parts of circles and cannot deal with straight lines of any length.

As pointed by Perona and Malik [15], the isotropic assumption is inappropriate for anisotropic elements. Instead, they introduced an *anisotropic assumption* and the anisotropic scale-space theory was developed. Related work includes [2, 18]. However, this series of work merely focused on image processing and did not go a step further towards telling what is the scale of edges.

In this paper, we study the scale of edges from the view

*This work was performed at Microsoft Research Asia.

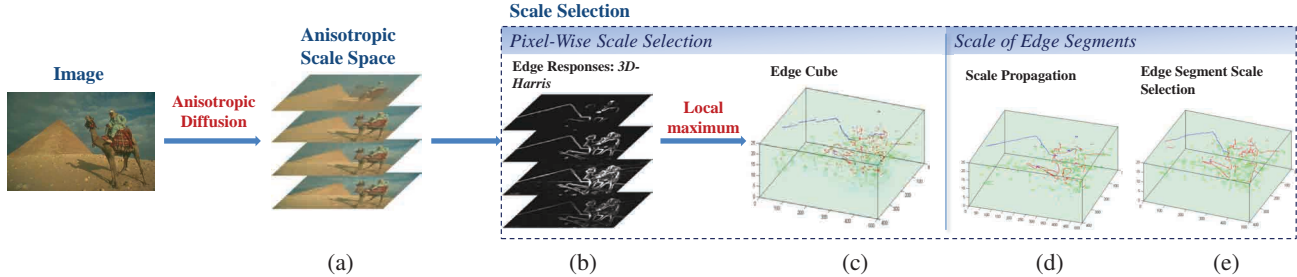


Figure 2. Flowchart of the technical details in calculating the scale of edges.

of human cognition. Fig. 1 shows two examples about how human beings observe edges in an image. In Fig. 1, the salient edges at different resolutions of an image are labeled by the subjects. These subjective examples illustrate the phenomenon that *images of different resolutions exhibit different level of edge details*. This could also be explained by the principle of human vision system: visual perception treats an image at several resolutions simultaneously, and human eye seems to possess visual information to “zoom in” on the right range of scale [7]. Based on these evidences, the scale of edges is formulated as a resolution-related parameter at which the edges are the most salient to human. Then, the existence of such optimal scale is studied in the anisotropic scale space. Moreover, we also explain and validate this process from the view of heat propagation. Finally, a quantitative definition of the scale of edges is given.

2) How to calculate the scale and what is it used for?

Based on the definition, an edge scale selection approach is proposed, whose flowchart is shown in Fig. 2. First, the scale space of an image is constructed via the anisotropic diffusion. Then, we extend the Harris detector [5] to a 3D version in the spatial-scale space to measure the edge response, and a novel scale selection algorithm is proposed to calculate the pixel-wise scales. As a result, an *edge cube* is constructed to indicate the positions and scales of edges. Finally, by introducing the assumption that edge pixels in the same edge structure are likely to have similar scales, two algorithms are adopted to refine the edge cube and calculate the scale of edge segments. Based on the data collected from a cognitive experiment, a comparison with Lindeberg’s method [10] is conducted to show the effectiveness of the proposed edge scale calculation approach.

Since the scale is a basic attribute of edges, better understanding of the scale of edges will benefit many edge-related applications, such as hierarchical edge parsing and boundary detection. As shown in the experiments, edges could be hierarchically represented: the edges with larger scales tend to be outer boundaries and the ones with finer scales are likely to be inner details. Besides, the importance of identifying the scale of edges is also illustrated in the boundary detection task. We not only propose a new low-level feature derived from the edge cube, but also illustrate how the scales influence the boundary detection process by proposing a scale-specific algorithm. Moreover, the proposed detector works well in localization owing to the anisotropic

assumption. This makes it suitable for applications which need high localization accuracy such as medical image processing and remote sensing image processing.

2. Preliminaries: The Anisotropic Scale Space

This section briefly reviews the anisotropic scale space. Suppose $I(x, y; t)$ represents the image intensity at position (x, y) and scale t . In contrast to the *homogeneity & isotropy* assumption in the isotropic scale space, Perona and Malik claimed that the diffusion should respect the existence of edges rather than blurring all the pixels in the same way [15]. Thus, they proposed the anisotropic diffusion and suggested a new way to construct the scale space. The basic idea is to perform more diffusion at non-edge positions and less on edges [1], where the prior knowledge of the existence of edges is indicated by the first-order gradient. The scale space was formulated by the anisotropic diffusion:

$$I_t = \text{div}(g(\|\nabla I\|)\nabla I), \quad (1)$$

where g is a monotonically decreasing function.

In practice, due to the lack of a closed-form solution, Eqn. 1 is usually approximated in an iterative way:

$$I^{t+1} = I^t + \lambda[g(\|\nabla I^t\|)\|\nabla I^t\|], \quad (2)$$

where the iteration number t is considered as a scale parameter [15] in the scale space.

3. The Scale of Edges

The study of the scale of edges starts from the observations from a group of human cognitive samples of observing edges at different resolutions, based on which a general description to the scale of edges is introduced. We further illustrate the existence of the optimal scales for edges, followed by a physical interpretation. Finally, a quantitative definition of the scale of edges is given.

3.1. What is the Scale of Edges?

To study the scale of edges, we first turn to human observations to the salient edges¹ at different scales of each image. As shown in Fig. 1, subjects were asked to label the most salient edges at each resolution. We can see that,

¹The saliency of edges is subjectively judged by human labelers.

as the image resolution decreases, some edges become unobservable while some others are still salient. For example, at low resolutions, the interior edges of the Sydney Opera seems invisible, whereas the back-shell structure is more perceivable at low resolutions than at higher resolutions.

Above observations show that an edge will only be perceivable in a range of resolutions, which is consistent with the study in [7]. We claim that the scale of edges is related to the saliency of edges observed by human, and this relationship results in that the scale of edges is determined by the perceivable range of resolutions. Therefore, we provide a general definition for the *scale of edges* as:

The scale of edges is a parameter to indicate at which resolution(s) an edge is the most salient for human.

The *scale space* was proposed to quantitatively study the image representations at different resolutions [7], with large scales modeling the images at small resolutions and vice versa. Thus, the *anisotropic scale space* [18] is naturally adopted for our analysis to the scale of edges due to its edge preserving property [15].

Inherited from the original definition in the scale space, the scale t in the anisotropic scale space might be a suitable parameter to describe the scale of edges when the image resolution changes. In the next part, we will further illustrate the existence of an optimal scale for edges in the scale space by analyzing the behaviors of edges.

3.2. Existence of Optimal Scales for Edges

Since the illuminance of any image will be flat at infinite large scales in the scale space [7, 15], which means all the pixels will be of the same grey value, thus we have

$$\lim_{t \rightarrow \infty} \|\nabla I(x, y; t)\| = 0, \forall (x, y) \in R^2. \quad (3)$$

As a result, the strength of edges which is measured by the first-order gradient is not always decreasing as the scale increases. Fig. 3 (a) shows the representations of a one-dimensional signal composed with a square signal overlaid by a sinusoidal distortion and Gaussian noises, filtered by different scales of anisotropic diffusion. Fig. 3 (b) illustrates the corresponding first-order derivatives. The edges in a one-dimensional signal are defined as the local extrema of the first-order derivatives. From Fig. 3 we can see that, different signals present different behaviors at different scales: from scale 1 to 2, the *sinusoidal* signal is enhanced, but it vanishes at scale 3; meanwhile, the *square* signal becomes more obvious from scale 2 to 3.

Based on the above observations, we can draw the following conclusions:

- 1) *The existence of the optimal scales:* For each edge component, its first-order gradient will increase within a certain range of scales as illustrated in Fig. 3, and then be suppressed to 0 if we continually increase the scale to infinitely large, as inferred from Eqn. 3. Therefore,

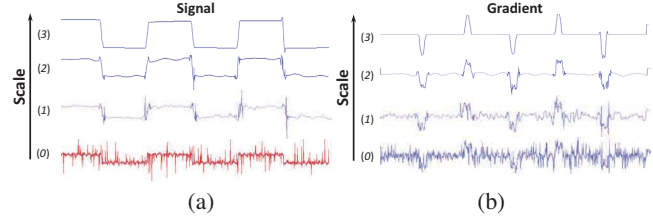


Figure 3. A square signal overlaid by a sinusoidal distortion and Gaussian noises, which is shown in red. The signals and their corresponding first-order derivatives at different scales after anisotropic diffusions are also shown. Note that the amplitudes are normalized to the same value for observing the details of signal changes.

for any edge there will always exist an optimal scale at which the response² $r(\nabla I(\cdot; t))$ reaches the maximum.

- 2) *Different level of edges are salient at different scales:* This can be inferred from the behaviors of different signals in Fig. 3. That is, different signals behave to be the most salient at different scales.

To conclude, above two evidences illustrate that there exist optimal scales for edges, with the ability of separating different signals at different scales. This optimal scale can be considered as the scale of edges. This is the foundation of scale selection and applications in the rest of this paper.

3.3. Physical Interpretation

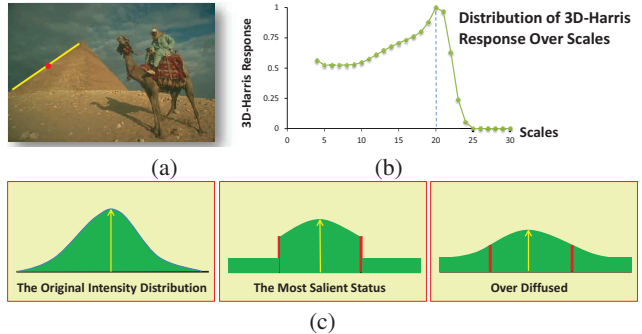


Figure 4. Illustration of the existence of the scale of edges from the point of view of heat propagation. The 3D-Harris edge responses of a given pixel at different scales are also shown in this figure.

An interpretation from the point of view of *heat propagation* is given in this section to further illustrate the existence of the scale of edges.

It is a common sense that the Gaussian filtering can be modeled as a heat propagation process in a uniform heat-transfer medium. The scale t is the time of propagation from the initial status to the current intensity distribution. Analogically, the representations of edges at different scales are also interpreted as a heat propagation process in this paper.

²More generally, we use $r(\cdot)$ to represent the edge response. The first-order gradient is the simplest case. In practice, $r(\cdot)$ can incorporate extra factors. E.g., in the 3D-Harris in Section 4.1, $r(\cdot)$ is the output of Eqn. 9.

However, the difference from the Gaussian filtering is that *the structure of edges changes the way that heat propagates*. The edges play a role of insulated containers in the heat propagation, making the trends inside and outside them very different. These insulated containers make the scale space anisotropic, and hence reinforce the edges at a certain range of scales. This process is visually demonstrated in Fig. 4 (c). As time goes on, there is one moment when the intensity inside a container is overflowed and the edge is over diffused. Thus, at this scale, the edge represented by the container has the largest gradient over scales.

A real example is also shown in Fig. 4 (a) and (b). To measure the edge response, the 3D-Harris response, which will be further explained in Section 4, is calculated at different scales for a fixed point (marked in red in Fig. 4 (a)). The distribution of the 3D-Harris responses of this point is shown in Fig. 4 (b). We can see that the edge response is first enhanced and then suppressed as the scale increases.

3.4. Quantitative Definition: the Scale of Edges

To conclude the above analysis and evidences from both human cognition and physical interpretation, we explicitly define the scale of edges as follows:

The scale of edges is the scale s in the anisotropic scale space satisfying that $r(\nabla I(\cdot; t))$ is a local maximum over scales, where $r(\cdot)$ is a monotonically increasing function and $r(\nabla I(\cdot; t))$ is the measurement of edge saliency, i.e. $s = \max_t r(\nabla I(\cdot; t))$.

This definition suggests the principle of scale selection for edges, which will be studied in the next section in detail.

4. Scale Selection for Edges

In this section, we study how to calculate the scale of edges. First, a novel 3D-Harris detector is proposed for calculating the scale of edge pixels. Then, two approaches, i.e. *Scale Propagation* and *Edge Segment Scale Selection*, are introduced to estimate the scale of edge segments by considering the structural information of edges.

4.1. Estimating the Pixel-Wise Scale of Edges

According to the definition of the scale of edges, the principle of scale selection for edge pixels is to find the local maxima:

$$s(x, y) = \arg \max_t r(\nabla I(x, y; t)), \quad (4)$$

where $r(\cdot)$ is a measurement of edge saliency.

To leverage the spatial/geometrical property of edges and avoid detecting isolated points, a 3D-Harris detector is proposed in the spatial-scale domain to measure the edge response of each pixel in the anisotropic scale space.

4.1.1 Harris Detector

The Harris detector [5] was proposed to detect corner points and edges by considering the change of image intensity

within a window when shifting around nearby regions. It evaluates the variations in two principal directions defined by the eigenvectors of the second-moment matrix:

$$M = g * \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix}, \quad (5)$$

and a *response* value for each pixel is calculated by:

$$R = \det(M) - k \cdot \text{trace}(M)^2 = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2, \quad (6)$$

where λ_i are the eigenvalues of M . By choosing a proper k , the pixels with negative response $R < 0$ are considered as edge candidates, since there exists a significant difference between the variances in two principal directions.

4.1.2 3D-Harris Detector for the Scale of Edge Pixels

In this work, we analyze the edges in the 3D scale space instead of in the 2D image plane. Therefore, a 3D-Harris detector is proposed in the scale space by introducing the following two constraints:

- 1) **Spatial constraint:** an edge should be local maxima of the spatial gradient over image intensity;
- 2) **Scale constraint:** the right scale of edges requires $\nabla I_t = \frac{\partial \|\nabla I\|}{\partial t} = 0$.

The first constraint is based on the local maxima of the gradient, whereas the other one is based on the zero-crossing of the gradient. The inconsistency between the two constraints makes it difficult to calculate the 3D version of the second-moment matrix M and the edge response R as in [5]. Thus, a transformed gradient $\nabla \tilde{I}_t$ is designed to convert the zero-crossing point $\nabla I_t = 0$ to a local maximum:

$$\nabla \tilde{I}_t = \text{sign}(\nabla I_t) \left(1 - \frac{\nabla I_t}{\max(\text{abs}(\nabla I_t))} \right). \quad (7)$$

Then, the second-moment matrix of 3D-Harris is defined:

$$M = g * \begin{pmatrix} I_x^2 & I_x I_y & I_x \nabla \tilde{I}_t \\ I_x I_y & I_y^2 & I_y \nabla \tilde{I}_t \\ I_x \nabla \tilde{I}_t & I_y \nabla \tilde{I}_t & \nabla \tilde{I}_t^2 \end{pmatrix}, \quad (8)$$

where the matrix M encodes the variances when shifting the 3D window in the direction of $[\Delta x, \Delta y, \Delta t]$, and the eigen-values of M measure the changes in the three principal directions.

Similar to the 2D situation, the response

$$R = \det(M) - k \cdot \text{trace}^3(M) = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3, \quad (9)$$

is defined to measure the significance of these changes.

Since edge points typically vary in less than two principal directions, only the points with negative responses will be considered as edge points, which is consistent with the

2D situation [5]. Moreover, as a larger contrast will lead to a larger $|R|$, $|R|$ is chosen as the measurement of edge saliency. Fig. 2 (a-b) shows an example, where (a) is the image representations at different scales and (b) is obtained by preserving the points with $R < 0$. The proposed 3D-Harris can detect the edges in the scale space, and the intensity $|R|$ reflects the strength of edges at different scales.

Therefore, using 3D-Harris to detect the scales of edge pixels is to find the optimal scale s to satisfy:

$$s(x, y) = \arg \max_t |R(x, y, t)|, R(x, y, t) < 0. \quad (10)$$

Fig. 2 (c) shows the detected edge pixels with their corresponding scales in the (x, y, s) -cube, which is called “*edge cube*” in this work³.

4.2. From Scale of Edge Pixels to Scale of Edges

The edges have structural information rather than isolated points. To determine the scale of edges, two structure-consistent assumptions are considered: 1) the pixels on an edge are spatially continuous, and 2) the pixels on the same edge segment usually have similar scales.

Thus, a two-step approach is introduced to estimate the scale of edges based on the edge cube. First, a greedy algorithm is used to extract edge segment. Then, two strategies, i.e. *Scale Propagation* and *Edge Segment Scale Selection*, are introduced to calculate the scale of edge segments.

4.2.1 Edge Segment Extraction

For each edge pixel $p_i = (x_i, y_i, s_i)$ in the edge cube, we use a greedy algorithm to find its adjacent edge pixel p_{i+1} on the same edge segment by considering two factors: the consistency in edge response $R(p) = R(x, y, s)$ and segment orientation $\vec{p_i p_{i+1}} = (x_{i+1} - x_i, y_{i+1} - y_i)$:

$$p_{i+1} = \arg \max_{p \in \mathcal{N}(p_i)} -|R(p) - R(p_i)| + \lambda \frac{\overrightarrow{p_{i-1} p_i} \cdot \overrightarrow{p_i p}}{|\overrightarrow{p_{i-1} p_i}| |\overrightarrow{p_i p}|}, \quad (11)$$

where $\mathcal{N}(p_i)$ is the spatial neighbors of p_i regardless of scale value, and the orientation change is measured by the cosine similarity. The point with the largest consistency is chosen as the successor of this segment. The algorithm begins from 1-neighbor points ($|\mathcal{N}(p_i)| = 1$), and stops if all non-isolated points ($|\mathcal{N}(p_i)| > 0$) are processed. Then, the edges are organized as a set of edge segments $\mathcal{C} = \{C\}$.

4.2.2 Scale Propagation

Let i ($i = 1, \dots, n$) denote the index of the i -th edge pixel on segment C , and s_i is the detected scale of the i -th edge pixel. We further refine the scales based on two assumptions: 1) nearby pixels on the same edge segment should

³In this paper, we use t to denote the scale in the scale space, and s to denote the selected optimal scale of edges.

have similar scales, and 2) the scales should be consistent with the results of 3D-Harris. Thus the cost function is:

$$\mathcal{F}(\mathcal{S}^*) = \frac{1}{2} \sum_{i,j=1}^n W_{ij} \|s_i^* - s_j^*\|^2 + \mu \sum_{i=1}^n \|s_i^* - s_i\|^2, \quad (12)$$

where s_i^* is the refined scale after scale propagation, and W_{ij} is the similarity to measure the geodesic distance between i and j . Then the scales $\mathcal{S}^* = \{s_i^*\}$ of edges pixels on edge C can be obtained by minimizing $\mathcal{F}(\mathcal{S}^*)$ [20].

For each edge segment C , the scale propagation algorithm is performed to smooth the scale and the edge response R . Then, a statistic of pixel-wise scales $S(C)$ such as mean or median can be used as the scale of C .

4.2.3 Edge Segment Scale Selection

Another approach to determine the scale of an edge segment C is to search the local maximum of its response $R(C, t)$ in the scale space, which is a robust statistic of the edge response of edge pixels on C , i.e. $R(C, t) = \text{stat}(\{R(x, y, t), \{x, y\} \in C\})$. Thus we have,

$$S(C) = \arg \max_t |R(C, t)|, R(C, t) < 0. \quad (13)$$

In this work, we use the median as the statistic for its robustness to outliers.

The edge cubes refined using the above two algorithms are shown in Fig. 2 (d-e). We can see that, the edge cube after scale propagation looks smoother than that after edge segment scale selection, since it provides continuous scales for edge pixels; and using edge segment scale selection, the scales of pixels on a same edge segment are identical.

Therefore, these two algorithms could be used in different scenarios. The scale propagation exhibits more information of an edge segment and might be more proper for depth estimation, object alignment, and affine estimation; whereas the edge segment scale selection is simpler for implementation and calculation. In the experiments of this paper, we adopt the edge segment scale selection strategy.

5. Experiments and Applications

In this section, we first evaluate the detected scales of edges. Then, as a natural application, the scale of edge is used in the task of boundary detection. Finally, we illustrate its potential for hierarchical edge parsing.

5.1. Evaluation: The Scale of Edges

A quantitative experiment is performed to evaluate the proposed definition and algorithms for the scale and scale selection of edges. We first introduce the Multi-Scale Edge Dataset which was collected by conducting a cognitive experiment, followed by the evaluation measurement. Then, the proposed algorithms are compared with Lindeberg’s method [10] for edge scale selection.

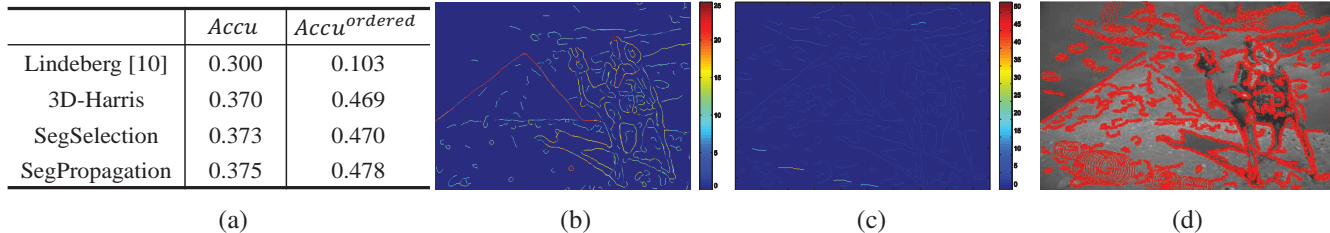


Figure 5. Comparison of different algorithms on the scale of edges. (a) Quantitative comparisons. (b) Heatmap of the scales detected by 3D-Harris. (c) Heatmap of the scales detected by Lindeberg’s method [10]. (d) Circlemap of the scales detected by Lindeberg’s method. Larger circles indicate the edge points of larger scales. For better viewing, please see the electronic version.

5.1.1 The Multi-Scale Edge Dataset

To address the problem of “what is the scale of edges from human cognition”, a perceptual experiment was performed to record the scale information of edges. We collected 40 photos rich in scale information but not too complex for manual labeling. Each image was shown to subjects at five resolutions⁴ from lowest to highest. For each resolution, a subject was first asked to label the edges which are visually salient enough to him / her. The edges labeled at lower resolutions were automatically shown at higher resolutions to reduce the labeling work. Examples are shown in Fig. 1.

5.1.2 Evaluation Measurement

Based on the labeled edges, we construct the groundtruth for evaluation. Naturally, an edge labeled at a lower resolution could be considered to have a larger scale than that labeled at a higher resolution. Let s_i^h and s_j^h denote the scales corresponding to a lower resolution and a higher resolutions at which edges C_i and C_j were labeled. Although there is no quantitative relationship between the image resolution and the scale of edges, we can have the following inference: $s_i^h > s_j^h$. We call s_i^h as the *groundtruth scale* of C_i . For each algorithm, if it can detect edge C_i , there will be an estimated scale s_i^d , which is called the *detected scale*. For each image, a fixed number of edge pixels, which were both labeled by subjects and detected by an algorithm⁵, were randomly sampled to compose the testing set.

Since we only know the non-quantitative relationship between the groundtruth scales, we need to evaluate whether the relationship between the detected scales for a pair of edge pixels is consistent with that between the groundtruth scales, and the qualified pairs are called *concordant pairs*. Thus, for each testing image I , the scale detection accuracies $Accuracy(I)$ and $Accuracy^{ordered}(I)$ are used as the measurements, in which $Accuracy(I)$ denotes the ratio of the number of *concordant pairs* to that of testing pairs, and $Accuracy^{ordered}(I)$ only considers the pairs with different *groundtruth scales* to avoid trivial solutions. In im-

⁴For the five resolutions, the length of the longer side of each image is 40, 80, 160, 320, and 640, respectively.

⁵As in the evaluation in boundary detection [14], if the distance between a detected point and the nearest groundtruth point smaller than a constant, i.e., 3 pixels, it will be considered as a correct detection.

plementation, 800 edge pixels were randomly sampled in each testing image, resulting $800 * 799/2$ edge point pairs. We repeat this sampling process for 10 times to stabilize the evaluation, and the average performance *Accuracy* and *Accuracy^{ordered}* ($Accu$ and $Accu^{ordered}$ for short) over all testing images are used as the evaluation measures.

5.1.3 Results and Comparisons

Fig. 5 (a) shows the comparison results of four algorithms: 3D-Harris, 3D-Harris with Scale Propagation / Edge Segment Scale Selection (*SegPropagation/SegSelection* for short), and Lindeberg’s method [10]. We can see that, all our methods greatly outperform Lindeberg’s method under both measurements. To be noticed, Lindeberg’s method performs rather bad when measured by $Accuracy^{ordered}$. This is because it wrongly assigned most pixels at different *groundtruth scales* to the same scales, and thus is less discriminative to the scale of edges.

More detailed comparisons are visualized in Fig. 5 (b-d). As shown in Fig. 5 (b), the scales calculated by 3D-Harris are able to reveal the structure of edges: the scale of the boundary of the pyramid is larger than the scales of the camel’s contour and shadow, which are further larger than the scales of the textures and the details inside the camel. On the other hand, we cannot see such a clear structure of edges in Lindeberg’s method, as shown in Fig. 5 (c). Fig. 5 (d) shows the scales detected by Lindeberg’s method in a more direct way. The size of red circles indicate the detected scales: larger circles correspond to the points of larger scales. Actually, the scale of Lindeberg’s method [10] measures the diffusion degree of a step edge, which can be viewed as the size of “shadow” region instead.

We can also see that, although the measurements are designed for evaluating pixel-wise scales, after adding the structure-consistent assumption, the performance of 3D-Harris is still slightly improved.

5.2. Application: Boundary Detection

Most of mainstream approaches for boundary detection such as [4, 13, 14, 16] are PB-based methods, abbr. *Probability of Boundary*, all of which learn a score for each pixel to indicate its probability to be a boundary point. The basic features used in these methods (named as PB-based fea-

Table 1. Comparison of different boundary detection algorithms.

Single Feature		Multiple Features	
Method	F	Method	F
CG	0.57	PB [14]	0.65
BG	0.59	Boosted [4]	0.66
TG	0.59	Multi-Scale PB [16]	0.68
RS	0.61	Scale-Specific PB	0.66

tures) include the Brightness Gradient (BG), Texture Gradient (TG), and Color Gradient (CG) [13, 14, 16].

In this section, we illustrate the effectiveness of the edge cube in boundary detection. First, a novel low-level feature derived from the edge cube is introduced (“RS” feature, short for “Response at a specific Scale”). Then, a modified PB [14] algorithm which naturally encodes the scale information is presented (“Scale-Specific PB”).

5.2.1 The RS Feature

Since the edge cube presents different levels of edge details at different scales, we simply project the edge cube vertically to a certain scale, and obtain a response map. The edge response R for each pixel is considered as the RS feature of this pixel. The RS feature is used in boundary detection in a similar way as existing low-level PB features [14]. The boundary detection results of different low-level features on the BSD300 dataset are shown in Table 1, in which the F-Score ($2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$) is used as the measurement. We can see that, the RS feature achieves a performance of F-Score = 0.61, which is higher than all other low-level features, such as TG (0.59), BG (0.59), and CG (0.57).

Besides the superior detection performance, the RS feature also has better localization accuracy in the detection task. When evaluating boundary detection algorithms [4, 13, 14, 16], a threshold d_t is used to tolerate the location shift. That is, if the distance between a detected point and its nearest groundtruth point is smaller than d_t , it will be considered as a correct detection. To compare the localization accuracy, we decrease d_t to make a stricter comparison. As shown in Table 2, the superiority of RS becomes more obvious as d_t becomes smaller.

5.2.2 Scale-Specific PB

In this part, we further investigate how the detected scales of edges benefit PB-based methods in boundary detection. Although previous work [16, 17] revealed the advantages of considering multiple scales in boundary detection, the relation between the scale and the existence of boundaries is not studied and how to use the scales is still unclear.

Different from existing work which extracts features for each pixel in a uniform scale [14] or simultaneously uses multiple scales [16], we propose to extract PB-based features at their corresponding detected scales resulted from the proposed scale estimation algorithm.

Table 2. Performance comparison of different features with threshold d_t decreasing. As d_t decreases, the superiority of RS over other features becomes more obvious, which shows the effectiveness of RS in edge localization.

Feature	d_t	0.0075 (default)	0.0025
	CG		0.57
TG		0.59 (+3.5%)	0.31 (-7.6%)
BG		0.59 (+3.5%)	0.41 (+20.7%)
RS		0.61 (+7.4%)	0.45 (+30.7%)

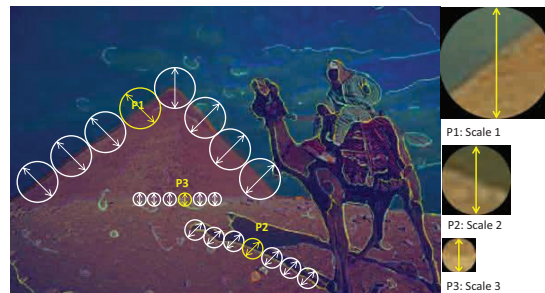


Figure 6. Scale-Specific PB: the features of each pixel is only extracted at the detected scale. In this example, three pixels of different scales are shown.

Given a testing image, we first detect the scales of all edge pixels, and then an image dilation operation is adopted to make scale detection more stable. For pixels of large scales, their PB-based features will be calculated within the disks of a large size, and vice versa. An example is shown in Fig. 6, in which the features of pixels at three different scale levels are extracted based on their corresponding scales (indicated by the disks with different sizes). More specifically, we uniformly divide the edge scales into 4 octaves, which correspond to the disk radiuses $(\sqrt{2})^{-1}$, 1, $(\sqrt{2})^1$, $(\sqrt{2})^2$ of the default value in [14]. For each pixel, we only extract its features on its corresponding scale. In the training stage, logistic-regression models are trained on every octave, and each model is only used to classify the pixels of the corresponding scale.

Comparison results of different algorithms are shown in Table 1. We can see that, the proposed Scale-Specific PB outperforms PB [14], but does not exceed Ren’s Multi-Scale PB [16], in which all the 6 octaves were used. However, the Multi-Scale PB is computationally expensive: the time cost will increase exponentially with the disk size. Instead, in our algorithm, only the right scale is used to reduce the computation cost. The Boosted Edge Learning [4] achieved the same performance as ours, which leveraged about 50,000 features for each pixel; whereas only four features and a linear classifier were adopted in the proposed algorithm.

This experiment is mainly to illustrate the rationality of the detected scale of edges, and its effectiveness in the boundary detection task, rather than design a complete boundary detection algorithm. There might be other smart ways to leverage the scale information to design a better boundary detection algorithm.

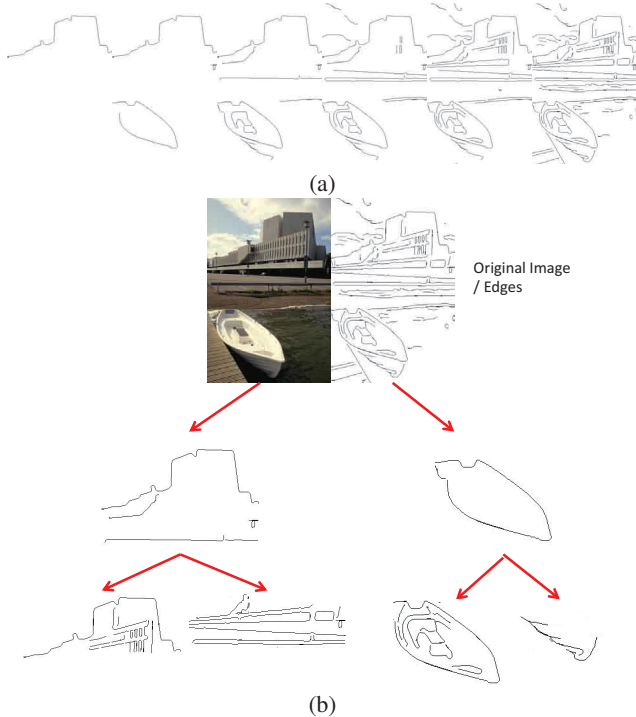


Figure 7. (a) The detected edges as the scale decreases. (b) An example of hierarchical edge parsing.

5.3. Application: Hierarchical Edge Parsing

As aforementioned, the edge segments of large scales are usually object boundaries, and as the scale decreases, more details will appear. This is further illustrated in Fig. 7 (a). The edge images were obtained by projecting the edge cube vertically to the planes of different scales. We can see that, in relatively large scales, the outer boundaries of the building and the boat are detected; as the scale decreases, more details include the windows of the building and the inside structures of the boat come out.

This shows the potential application of hierarchical edge parsing, with an example shown in Fig. 7 (b). The edges with the largest scales can be first clustered into objects, based on the observation that most boundaries have large scales and different objects are separated. By defining the spatial relationships including “containing”, “continuous” and “adjacent”, edge details with smaller scales can gradually appear to enrich each object. This process is demonstrated in Fig. 7 (b). We believe the technology of edge parsing will benefit edge-based image indexing and sketch-based image retrieval [3]. Object parsing is out of the scope of this paper, and more complex factors should be considered to design a practical algorithm. Thus, we only show the potential of the scale of edges in solving this problem.

6. Conclusions

In this paper, we have studied the scale of edges, a very fundamental problem in computer vision. Based on human perception and physical interpretation, a quantitative defini-

tion for the scale of edges was provided, based on which a scale detection algorithm was proposed. The Multi-Scale Edge Dataset was collected to evaluate the rationality of the detected scales. Moreover, we also illustrated the effectiveness of identifying the scale of edges in potential applications such as boundary detection and hierarchical edge parsing. We hope this work can inspire more studies on edge-related work such as feature detection, representation, and matching.

Acknowledgement

Xian-Ming Liu and Hongxun Yao were supported by National Natural Science Foundation of China (Grant No. 61071180) and Key Program (Grant No. 61133003).

References

- [1] L. Alvarez, P.-L. Lions, and J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. ii. *Siam Journal on Numerical Analysis*, 1992.
- [2] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *TIP*, 1998.
- [3] Y. Cao, C. Wang, L. Zhang, and L. Zhang. Edgel index for large-scale sketch-based image search. In *CVPR*, 2011.
- [4] P. Dollár, Z. Tu, and S. Belongie. Supervised learning of edges and object boundaries. In *CVPR*, 2006.
- [5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, 1988.
- [6] F. Jurie and C. Schmid. Scale-invariant shape features for recognition of object categories. In *CVPR*, 2004.
- [7] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 1984.
- [8] T. Lindeberg. On scale selection for differential operators. In *Proc. 8th Scandinavian Conf. on Image Analysis*, 1993.
- [9] T. Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 1994.
- [10] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *IJCV*, 1998.
- [11] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 1998.
- [12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.
- [13] M. Maire, P. Arbeláez, C. Fowlkes, and J. Malik. Using contours to detect and localize junctions in natural images. In *CVPR*, 2008.
- [14] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 2004.
- [15] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *PAMI*, 1990.
- [16] X. Ren. Multi-scale improves boundary detection in natural images. *ECCV*, 2008.
- [17] X. Ren and J. Malik. A probabilistic multi-scale model for contour completion based on image statistics. *ECCV*, 2002.
- [18] J. Weickert. *Anisotropic diffusion in image processing*. 1998.
- [19] A. P. Witkin. Scale-space filtering. In *IJCAI*, 1983.
- [20] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS*, 2004.