

# TransHist: Occlusion-robust shape detection in cluttered images

Chu Han<sup>1</sup>, Xueting Liu<sup>1</sup>, Lok Tsun Sinn<sup>1</sup>, and Tien-Tsin Wong<sup>1</sup> (✉)

© The Author(s) 2018. This article is published with open access at Springerlink.com

**Abstract** Shape matching plays an important role in various computer vision and graphics applications such as shape retrieval, object detection, image editing, image retrieval, etc. However, detecting shapes in cluttered images is still quite challenging due to the incomplete edges and changing perspective. In this paper, we propose a novel approach that can efficiently identify a queried shape in a cluttered image. The core idea is to acquire the transformation from the queried shape to the cluttered image by summarising all point-to-point transformations between the queried shape and the image. To do so, we adopt a point-based shape descriptor, the *pyramid of arc-length descriptor* (PAD), to identify point pairs between the queried shape and the image having similar local shapes. We further calculate the transformations between the identified point pairs based on PAD. Finally, we summarise all transformations in a 4D transformation histogram and search for the main cluster. Our method can handle both closed shapes and open curves, and is resistant to partial occlusions. Experiments show that our method can robustly detect shapes in images in the presence of partial occlusions, fragile edges, and cluttered backgrounds.

**Keywords** shape matching; shape detection; transformation histogram

## 1 Introduction

Shape matching plays an important role in various computer vision and graphics applications such as shape retrieval, object detection, image editing,

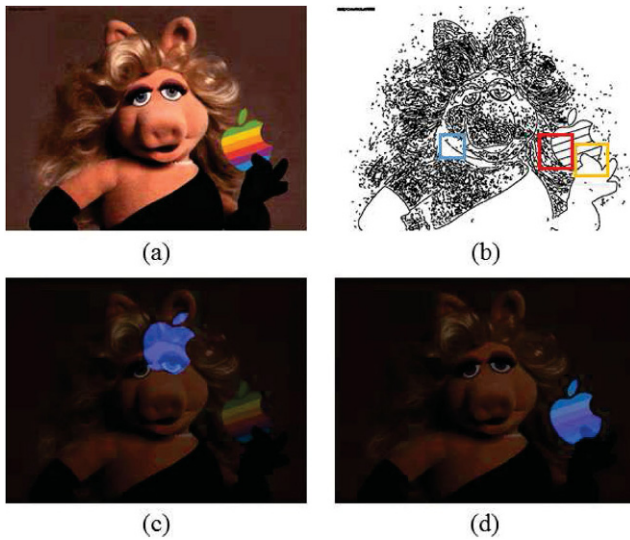
image retrieval, etc. Compared to gradient and texture features, shape features are much more reliable when objects are characterized by distinctive shapes, such as road signs in images and videos. In this paper, we focus on detecting shapes in cluttered images by analyzing point-to-point transformations.

In the early days, methods were proposed to measure shape similarity by transforming shapes into other domains, using, e.g., wavelet-based transforms [1] and Fourier transforms [2–4]. Methods have also been proposed to transform shapes into the curvature domain, using curvature inflection points for shape matching [5, 6]. Later, various shape descriptors were proposed and used for measuring shape similarity. Shapes may also be described using triangle areas, by forming a set of triangles at a reference point [7]. Use has also been made of an integral kernel to extract shape characteristics within a region centered at a reference point [8, 9]. The state-of-the-art *shape context* shape descriptor utilizes a log-polar diagram to statistically record the spatial distribution of shapes at each sample point [10].

However, most existing shape matching algorithms or shape descriptors are only designed for matching shapes with clean and clear edges. Unsatisfactory results may be obtained if they are used to detect shapes in cluttered images. First of all, edges in cluttered images are fragile edges: shapes may be cut into fragments (blue box in Fig. 1(b)). With no tolerance for fragile edges, whole-shape descriptors naturally fail to detect shapes with fragile edges. Furthermore, region-based shape descriptors are also unable to extract features from fragile edges since they assume that shapes are closed. Secondly, partial occlusions (yellow box in Fig. 1(b)) typically occur, again hindering whole-shape descriptors from finding

<sup>1</sup> The Chinese University of Hong Kong, Hong Kong, China. E-mail: C. Han, [chan@cse.cuhk.edu.hk](mailto:chan@cse.cuhk.edu.hk); X. Liu, [xtliu@cse.cuhk.edu.hk](mailto:xtliu@cse.cuhk.edu.hk); L. T. Sinn, [lbsinn@cse.cuhk.edu.hk](mailto:lbsinn@cse.cuhk.edu.hk); T.-T. Wong, [ttwong@cse.cuhk.edu.hk](mailto:ttwong@cse.cuhk.edu.hk) (✉).

Manuscript received: 2017-12-29; accepted: 2017-12-31



**Fig. 1** Challenges in shape detection in cluttered images. (a) Input image. (b) Edge map of (a). (c) PAD with distance field. (d) Our result.

the correct solution. Lastly, the background in cluttered images may be extremely noisy (red box in Fig. 1(b)). Noise greatly affects shape descriptors in the spatial domain. Moreover, cluttered backgrounds also increase computational costs.

The *inner-distance shape context* [11] extended the original shape context to tackle partial-shape matching, but it still requires the shapes to be closed. Recently, Kwan et al. [12] proposed a point-based shape descriptor called the *pyramid of arc-length descriptor* (PAD) to tackle partial occlusion, and their method is also applicable to open curves. However, their method is greatly affected by cluttered backgrounds since they rely on distance fields for final shape matching. Moreover, it is computationally expensive.

In this paper, we present an efficient shape detection method tailored for shape detection in cluttered images. We found that, while shape detection can be regarded as finding the optimal transformation from a queried shape to the shape in the image, we may first identify the optimal transformation for each point on the shape and then summarise all transformations to get the optimal transformation for the whole shape. In order to find the optimal transformation for each point efficiently, we adopt PAD [12] as the shape descriptor for each point and find the optimal corresponding point in the image by comparing similarity of

PAD. We use PAD because of its scale, rotational, and translational invariance. Furthermore, PAD is applicable to both closed and open curves, and is robust to partial occlusion. We then calculate the transformation between corresponding points based on PAD (Fig. 5), and form a 4D transformation histogram to summarise the transformations of all points on the shape. The main cluster of transformations is the detected result. Figure 1(c) shows the result detected by directly applying PAD, while Fig. 1(d) shows the result detected by our method. It can be seen that our method is more robust to noise and therefore more appropriate for detecting shapes in cluttered images. We later provide the results of two experiments to validate the robustness of our method to partial occlusion and cluttered backgrounds.

Our contributions can be summarized as follows:

- an efficient shape detection method for detecting shapes in cluttered images;
- which is well able to simultaneously handle fragile edges, partial occlusion, and cluttered backgrounds.

## 2 Related work

Detecting shapes in images is a fundamental problem in computer vision and graphics. Over the past few decades, various two-dimensional shape descriptors have been proposed to describe the characteristics of a shape. They can be broadly classified into two categories: global shape descriptors, which describe the characteristics of the whole shape, and local shape descriptors, which describe the local characteristics of a shape at specific points.

### 2.1 Global shape descriptors

A popular family of global shape descriptors performs domain transform on the shape. Fourier descriptors [2–4] can be used as shape signatures that capture shape characteristics (e.g., centroid distance and cumulative angles along the shape boundary) in the Fourier domain. Wavelet descriptors [13] rely on wavelet transforms to obtain a multi-resolution representation of a shape from the shape boundary. Radon descriptors [14] rely on R-transforms, a variant of Radon transforms, of the shape to obtain the shape properties. Curvature scale space (CSS) [5, 6] describes a shape by

recording its inflection points at different level of smoothing. It represents the changes in location of inflection points on the shape boundary with smoothing. Later, Lee et al. [15] proposed shape signature harmonic embedding (SSHE), which uses discrete harmonic functions to replace smoothing in the construction. Descriptors have also been proposed based on moment theory, including Hu moments [16], Zernike moments [17], and image moments [18, 19]. Bernier and Landry [20] proposed a polar representation that plots the orientation of each boundary point referenced from the centroid of the shape as a description.

However, all of the above global shape descriptors describe shapes in an integral manner. They do not extract any local detail of the shape. Thus, we cannot directly use these global shape descriptors to measure shape similarity for open curves or complex shapes. Furthermore, global shape descriptors may also fail when partial occlusions exist.

## 2.2 Local shape descriptors

Local shape descriptors are generally point-based, describing local shape characteristics at reference points. One descriptor is built for each reference point, and all descriptors together form a rich description of the whole shape.

Shape context [10] is the state-of-the-art point-based shape descriptor. It describes the shape distribution of boundary points by a log-polar diagram centered at a reference point. Mori et al. [21–23] introduced fast pruning to speed up the matching process for shape contexts. Inner distance shape contexts [11] are an extension of shape contexts which use inner distance instead of traditional Euclidean distance in length measurement. However, shapes need to be closed to measure inner distances. Furthermore, all methods based on shape contexts are error-prone when used to detect shapes in cluttered images since shape contexts describe shapes in raster space.

Tănase et al. [24, 25] proposed use of a turning function that measures the turning angle along the boundary of the shape. Triangle area representation (TAR) [7] describes the shape of a reference point by the areas of triangles starting from this point. Integral invariants [8] describe the shape of a reference point using the integral kernel within a region centered at this point. Hong and

Soatto [9] further proposed a multi-scale integral invariant approach.

Although these point-based shape descriptors describe local shape characteristics, global normalization is still needed to achieve scale invariance, as they are not inherently scale invariant. These point-based descriptors cannot be directly applied to detecting shapes in cluttered images.

Recently, Kwan et al. [12] proposed a locally scale-invariant shape descriptor, the pyramid of arc-length descriptor (PAD). The locally scale-invariant property enables more robust detection of shapes in the presence of cluttered edges and noise. However, PAD only represents a very limited range of local shapes at a reference point, so further evaluation is needed for shape matching. While Kwan et al. used a distance field for this purpose, the resulting ability to detect shapes in cluttered images is poor, since distance fields are quite sensitive to noise in raster images. Instead, we adopt a transformation histogram for shape matching, making our method much more robust to noise.

## 2.3 Shape-based object detection

Various methods have been proposed to detect objects in 2D or 3D space [26, 27]. Here, we only focus on object detection in 2D, i.e., images. Most existing methods are based on shape context [10] since it is the state-of-the-art shape descriptor. In particular, Lian et al. [28] proposed to detect shapes with a novel outlier-resistant shape context distance that ignores outliers in norm-2 distance in the original shape context. Thayanathan et al. [29] introduced a continuity constraint into shape context matching, restricting the correspondence to be formed by nearby points. However, these methods cannot achieve scale invariance and are only applicable to detecting shapes at the same scale.

Riemenschneider et al. [30] proposed a partial-shape matching method to locate objects. However, their method is not truly scale invariant. Shape bands [31] can tolerate shape deformation to some extent, within a fixed bandwidth. But this method is also not scale invariant and is error-prone in the presence of cluttered edges. Cheng et al. [32] proposed to use a boundary band map to search for repeated elements with similar shapes, but user interaction is needed. The chordigram method [33] works by first forming a set of chord

by joining boundary points together, and then uses lengths, orientations, and normal directions of the boundary points forming the chords to form a chord diagram for shape matching. However, this method is not scale invariant and rotational invariant. Chi and Leung [34] proposed to decompose a shape into primitives and perform partial-shape matching by searching an indexed structure of primitives. However, they do not take scale invariance into consideration. In contrast, our approach achieves scale invariance and rotational invariance; only a single description is needed to describe the local shape in a scale-invariant and rotational-invariant manner.

Methods based on neural networks have also been proposed to detect objects [35–37]. While neural networks may achieve better results than traditional low-level methods, they rely highly on training data. In this paper, we aim to detect shapes in cluttered images by only relying on the shapes' characteristics.

### 3 System overview

In this paper, we propose a novel shape detection method which calculates and analyzes a transformation histogram relating the queried shape to the cluttered image. Figure 2 shows the framework of our method. Given a cluttered input image and a queried shape, we first extract the edges from both the image and the queried shape. To extract edges from the cluttered image, we use the Canny edge detector [38]. For the queried shape, we simply identify all boundary pixels as edges of the shape. We then calculate the local shape feature of each point using an existing point-based shape descriptor, the pyramid of arc-length descriptor (PAD) [12]. Note that PAD can be computed for all points no matter whether they are on closed or open curves. Moreover, PAD

only describes local shape features along a single edge. No redundant or disturbed information from the cluttered background is embedded in PAD. In addition, it is scale invariant, rotational invariant, and translational invariant, allowing the detection of shapes with changes in size and orientation.

The next step is to find all pairs of edge points with similar PAD features (see Section 4.1). We observe that there is a high probability that the points will be correctly matched, and the transformations for all correctly matched point pairs should be quite similar (Fig. 6). Thus, we first calculate the transformation for each pair of edge points in correspondence (see Section 4.2), and then form a transformation histogram using all transformations, to identify the main cluster that represents the transformation of the whole shape (see Section 4.3).

Our method is fully parallelisable for use on a GPU.

## 4 Shape detection in cluttered images

### 4.1 Point-to-point matching via PAD

Given the edge maps from an image and a query shape, we first extract the shape features of all reference points on the edges by a local point-based shape descriptor, pyramid of arc-length descriptor (PAD) [12]. The features extracted from PAD is locally scale invariant, rotational invariant, and translational invariant. It can provide the precise transformation of two points with similar local shapes, which perfectly matches our requirement. But we still want to point out that our framework actually can accept all local shape descriptors that provide point-to-point transformations.

For the completeness of the paper, we briefly introduce PAD and how it is used for local shape matching between two reference points in the

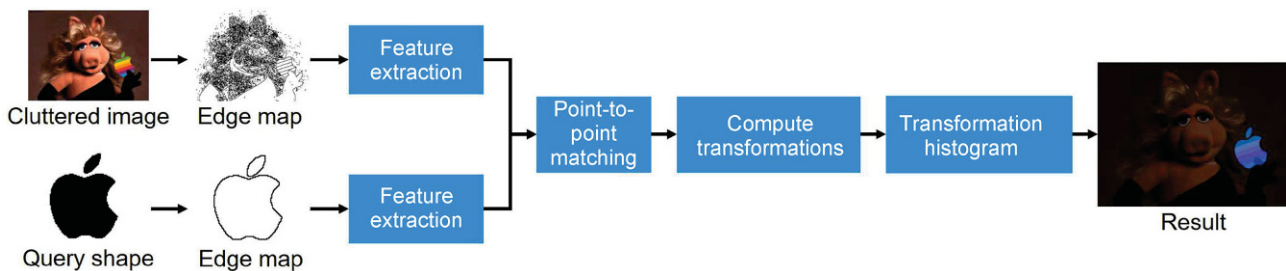


Fig. 2 Framework of our approach.

following. The PAD shape feature is extracted from the integral of absolute curvature (IAC) [39]. Given a curve (Fig. 3), the integral of absolute curvature  $\tau$  over a curve segment between points  $s$  and  $t$  is defined as

$$\tau(s:t) = \int_s^t |\kappa(x)| dx \quad (1)$$

where  $\kappa(x)$  is the curvature at point  $x$  on the curve.

PAD encodes the shape by combining the IAC domain and the arc-length domain. It constructs a pyramid of arc-length intervals centered at a point  $p$ , such that each interval corresponds to a fixed integral value of absolute curvature ( $2^i \Delta\tau$ ), cumulated from  $p$ . The pyramid of arc-length  $L_i$  and  $R_i$  can be extracted by integrating different levels of absolute curvature in the IAC domain. For level  $i$ , the two intervals  $(l_i : p)$  and  $(p : r_i)$  have the same IAC value  $2^i \Delta\tau$ , so

$$\tau(l_i : p) = \tau(p : r_i) = 2^i \Delta\tau \quad (2)$$

After cumulating arc-length on both left and right hand sides at different levels, the PAD vector is defined by using this set of arc-length values. The corresponding arc-lengths of these IAC intervals form the initial PAD vector  $\mathbf{M}_{\text{init}}$ . Figure 4 shows the set of intervals sampled for 5 levels and the IAC value accumulated for intervals from each level.

$$\mathbf{M}_{\text{init}}(p) = \begin{bmatrix} L_{n-1} & L_{n-2} & \cdots & L_0 \\ R_0 & R_1 & \cdots & R_{n-1} \end{bmatrix}^T \quad (3)$$

Normalization is performed to achieve scale invariance:

$$\mathbf{M}_n(p) = \begin{bmatrix} \frac{L_{n-1}}{2^{n-1}L_{n-1}} & \frac{L_{n-2}}{2^{n-2}L_{n-1}} & \cdots & \frac{L_0}{2^0L_{n-1}} \\ \frac{R_0}{2^0R_{n-1}} & \frac{R_1}{2^1R_{n-1}} & \cdots & \frac{R_{n-1}}{2^{n-1}R_{n-1}} \end{bmatrix}^T \quad (4)$$

The final PAD vector  $\mathbf{m}(p)$  is defined as

$$\mathbf{m}(p) = s \begin{bmatrix} \hat{l}_{n-2} & \hat{l}_{n-3} & \cdots & \hat{l}_0 \\ \hat{r}_0 & \hat{r}_1 & \cdots & \hat{r}_{n-2} \end{bmatrix}^T \quad (5)$$

where  $\hat{l}_i = \frac{L_i}{2^i L_{n-1}}$ ,  $\hat{r}_i = \frac{R_i}{2^i R_{n-1}}$ , and  $s \in \{+1, -1\}$ .

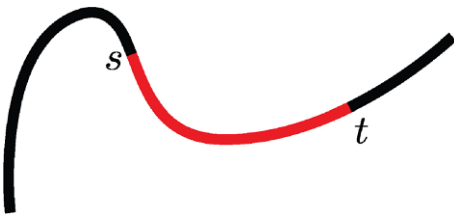


Fig. 3 An interval on a curve.

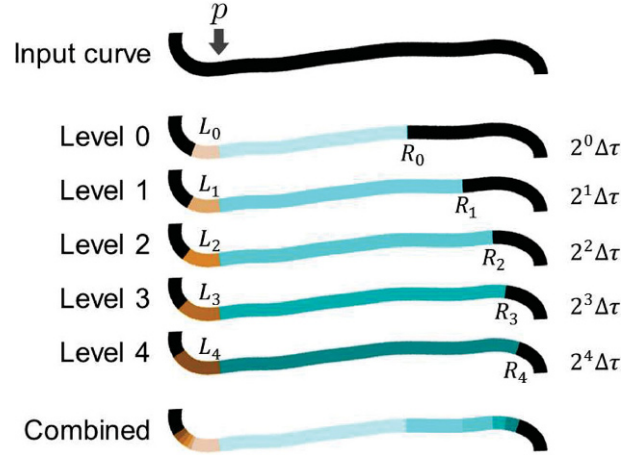


Fig. 4 Intervals sampled with multiple levels of PAD.

Here,  $s \in \{+1, -1\}$  is the sign of curvature at  $p$ , indicating whether the curve is convex (+1) or concave (-1) near the point of interest.

The PAD distance (similarity) between two local shapes around two points  $p$  and  $q$  is denoted  $D_{p,q}$ , and is the  $l^\infty$ -norm distance of the difference of two PAD vectors:

$$D_{p,q} = \|\mathbf{m}(p) - \mathbf{m}(q)\|_\infty \quad (6)$$

We can now estimate the local shape similarity of two points using the defined PAD similarity. We find all point pairs (one on the queried shape and the other on the image) with PAD distance larger than or equal to  $K = 0.2$  and denote these point pairs as *matching pairs*. We may increase  $K$  to enforce the matching pairs to have more similar local shapes, but with reduced tolerance for shape deformations.

#### 4.2 Transformation of matching pairs

We define the transformation between two points, and thus two local shapes, to be a 2D transformation comprising scaling, rotation, and translation along  $x$ - and  $y$ -axes. For example, Fig. 5 shows a queried shape  $Q$  (left) and an edge map  $E$  (right) with an identified matching pair  $p_i \in Q$  and  $q_j \in E$ . We build a vector (red arrows in Fig. 5) of the endpoints of the last level coverage of PAD for points  $p_i$  and  $q_i$ . The relative magnitude of the two vectors indicates the change in scale  $s_{p_i,q_j}$  between the two shapes under the PAD coverage. The angular difference of the two vectors indicates the change in orientation  $\theta_{p_i,q_j}$  of the two shapes locally. The translation between the two points  $(x_{p_i,q_j}, y_{p_i,q_j})$  can be computed as the spatial distance between the two

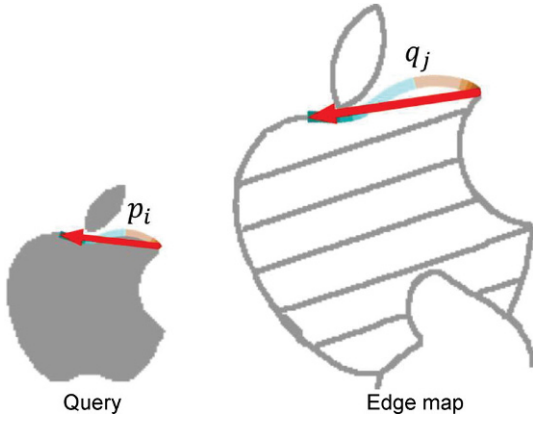


Fig. 5 Vector formed by last level PAD coverage.

vectors. Then we can write the transformation  $T_{i,j}$  as

$$T_{i,j} = [s_{i,j}, \theta_{i,j}, x_{i,j}, y_{i,j}] \quad (7)$$

where  $i$  and  $j$  are indices of points on the queried shape  $Q$  and the edge map of the cluttered image  $E$  respectively.

Note that the transformation model used here first translates the queried shape to the position defined by the vector, and then scales and rotates the queried shape correspondingly. Translation is referenced to the centroid of the shape in order to avoid deviations in position of the matching pair. This avoids translations being affected by the scaling and rotation components.

### 4.3 Transformation histogram

We now obtain a set of transformations  $\mathcal{T} = \{T_{i,j}\}$  for all matching pairs. All these transformations hint at possible locations of the queried shape in the cluttered image. It can be easily observed from Fig. 6 that matching pairs between two similar shapes should have similar transformations. Based on this observation, we use a transformation histogram to cluster the transformations.

Before putting these transformations into the histogram, we first normalize each component for better comparison. Given the original scale  $s_{i,j}$  for a transformation  $T_{i,j}$ , we assume the scale of the queried shape cannot be larger than the size of the cluttered image. Therefore, we normalize the scale by the diagonal of the cluttered image. We only take the normalized scale  $s'_{i,j} \in [0, 1]$  as an effective transformation and discard all transformations with normalized scale larger than 1. For orientation,

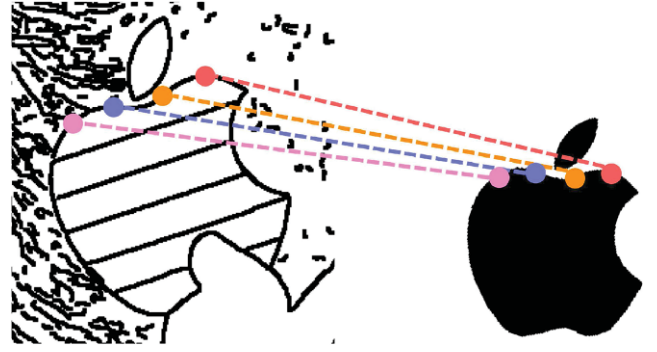


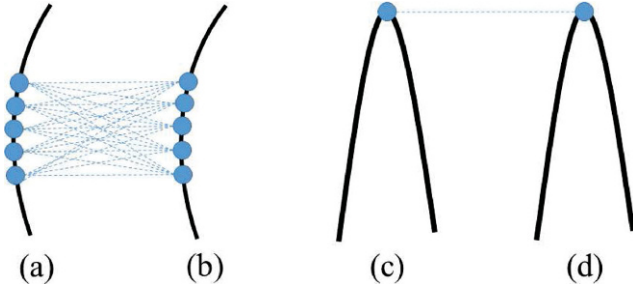
Fig. 6 Matching pairs along similar boundary portions. Each differently colored point is a PAD point involved in a different matching pair. Dashed lines show correspondences between matching PAD points.

we simply normalize it by the maximum possible rotation, i.e.,  $\pi$ . The range of the normalized rotation  $\theta'_{i,j}$  is thus  $[-1, +1]$ . We also normalize the  $x$ -translation  $x_{i,j}$  and the  $y$ -translation  $y_{i,j}$  by the width and height of the cluttered image respectively. Any normalized translation components  $x'_{i,j}$  and  $y'_{i,j}$  outside the range  $[0, 1]$  are taken as unsuitable transformations and discarded.

Each matching pair contributes a score to its corresponding histogram bin. We define the score to be

$$s_{i,j} = \frac{\kappa(i)\kappa(j)}{1 + D(i,j)} \quad (8)$$

where  $D(i,j)$  is the PAD distance of  $i$  and  $j$ , and  $\kappa$  denotes curvature. The aim is to consider local similarity and local smoothness for each matching pair  $(i,j)$ . We want locally more similar matching pairs to contribute higher scores. Since a smaller PAD distance means shapes are more similar, we thus put  $D(i,j)$  in the denominator. We further weight the score by local smoothness of the matching pair  $(i,j)$ , as smoother edges deliver less shape information and are more likely to be matched with other smooth edges. On the contrary, more rapidly changing edges contain more information. Figure 7 shows examples of matching two smooth edge segments and two more rapidly changing edge segments. Dashed lines indicate pairs of points which are locally similar and form a matching pair. We can easily observe that each point in Fig. 7(a) matches several locally similar points in Fig. 7(b). In contrast, the indicated point in Fig. 7(c) only matches one point in Fig. 7(d). Without consideration of local smoothness, the



**Fig. 7** Locally similar matching pairs with similar transformations. (a) and (b) are two smoother edges. (c) and (d) are two more rapidly changing edges. Dashed lines join locally similar pointers. Smoother edges have many more correspondences than more rapidly changing edges.

transformation histogram will be dominated by matching pairs of smooth points. We overcome this issue by weighting scores of matching pairs by their curvatures,  $\kappa(i)$  and  $\kappa(j)$ .

For each bin  $n$  corresponding to a certain transformation range  $[T_n, T_{n+1})$ , the final score  $S_n$  is the sum of scores of all matching pairs for that bin:

$$S_n = \sum_{i,j} s_{i,j}, \quad T_{i,j} \subseteq [T_n, T_{n+1}) \quad (9)$$

In practice, we set the numbers of bins for scale, rotation, and  $x$ - and  $y$ -translations to  $\{10, 10, 50, 50\}$ . By finding the bin with the largest score in the histogram, we can find the target shape in the image. Since it is possible that no similar shape exists in the cluttered image, we set a threshold on the fraction of the points for a match to exist. If the fraction is too low (empirically  $< 30\%$ ), we conclude that there is no similar shape in the image and return no match.

## 5 Experiments

In this section, we describe several experiments conducted to evaluate our shape detection method. Firstly we show some results of detecting shapes in cluttered images. Note that all results are directly plotted onto edge maps to aid visualization. We conduct two further experiments to validate the robustness to occlusion and cluttered edges respectively.

In these experiments, we compare our method to 6 existing shape matching methods, including PAD using a distance field [12], shape contexts [10], inner-distance shape contexts [11], curvature scale space [6], triangle area representation [7], and

integral invariants [8], abbreviated as PAD-DF, SC, IDSC, CSS, TAR, and II, respectively.

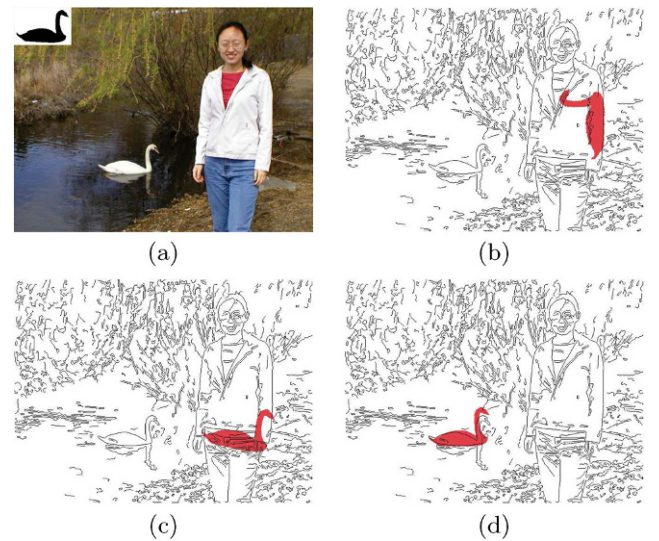
### 5.1 Shape detection in cluttered image

First we show a few results to demonstrate the ability of our method to detect shapes in cluttered images.

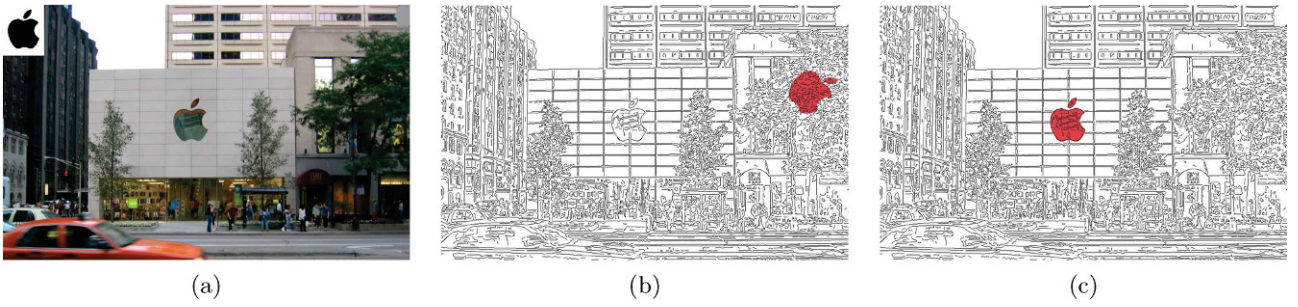
Figure 8(a) shows a real life photo containing a swan in the background. Since most existing shape descriptors cannot support open curves, we only compare our method with PAD-DF [12] and IDSC [11]. IDSC fails to detect the swan as inner distance is not defined for points across the disjoint components (see Fig. 8(b)). PAD-DF fails to locate the swan as it is confused by the clutter (see Fig. 8(c)). Our method successfully finds the location of the swan in the image (see Fig. 8(d)). Since it is unfair to compare IDSC in such cases, in the next two results, we only compare our method with PAD-DF.

Figure 9(a) shows an image from an object detection dataset, the ETHZ shape dataset [40]. The aim is to detect the Apple logo (shown at top left in Fig. 9(a)). Leaves on trees in the right part of the image lead to crowded edges (see Fig. 9(b)) which leads to incorrect detection results for PAD-DF (the Apple logo marked in red). In contrast, our method successfully avoids the effects of crowded edges and correctly matches the Apple logo in the image (see Fig. 9(c)).

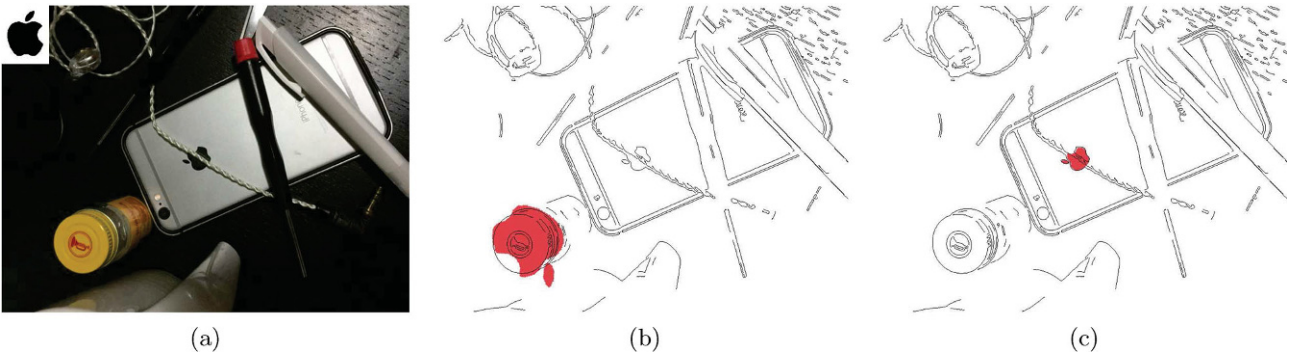
We show another example in Fig. 10(a) where the Apple logo is partially occluded by a wire.



**Fig. 8** Matching a swan. (a) Original image with query at top left. (b) IDSC result. (c) PAD-DF result. (d) TransHist result.



**Fig. 9** Matching an Apple logo on a window. (a) Original image with query at top left. (b) PAD-DF result. (c) TransHist result.



**Fig. 10** Matching an occluded Apple logo. (a) Original image with query at top left. (b) PAD-DF result. (c) TransHist result.

Figure 10(b) shows the result generated by PAD-DF. When detecting shapes in real images, PAD-DF is much more error-prone in the presence of crowded edges. Even with partial occlusion, our method can still survive and recognize the Apple logo correctly (see Fig. 10(c)).

## 5.2 Tolerance to occlusion

Partial occlusion is one of the major challenges in shape detection. However, this phenomenon is quite common in real world scenarios. This experiment aimed to explore the tolerance to partial occlusion, using the dataset proposed by Kwan et al. [12]. It contains a set of shapes and clipped instances of them. The shapes are from the MPEG7 dataset [41], and comprise 20 classes each with 70 shapes. Each shape is gradually occluded from left to right. The occlusion rate goes from 0% to 90% in 10% increments, giving 14,000 shapes and partial shapes in total.

We follow the rendering approach used by Kwan et al. [12] to quantify the goodness of matching: we render the clipped instance (after transformation) on top of the original shape. Let  $C$  be the set of pixels

belonging to the transformed clipped instance in the space of the original shape and  $C_o$  be the set of pixels where  $C$  should be, again in the space of the original shape. We then measure the matched fraction  $\gamma$  as follows:

$$\gamma = \frac{|C \cap C_o|}{|C \cup C_o|}$$

$\gamma = 1$  indicates a perfect match. Due to rasterization and numerical errors, we may not get a perfect match even if the match is visually perfect. Hence, we regard a transformation with  $\gamma > 0.95$  as a successful match.

Figure 11 plots the success rate against the degree of occlusion. We can see that PAD-DF and our method can still recognize shapes even in the presence of significant occlusion: even with 80% occlusion, we still achieve a success rate of around 25%. Most other descriptors, including TAR, SC, CSS, and II, are unable to deal with partial occlusion, and their success rates drop extremely quickly. Clearly, whole shape matching descriptors are inappropriate for measuring shape similarity in the presence of occlusion. Our transformation approach also outperforms PAD-DF, since the

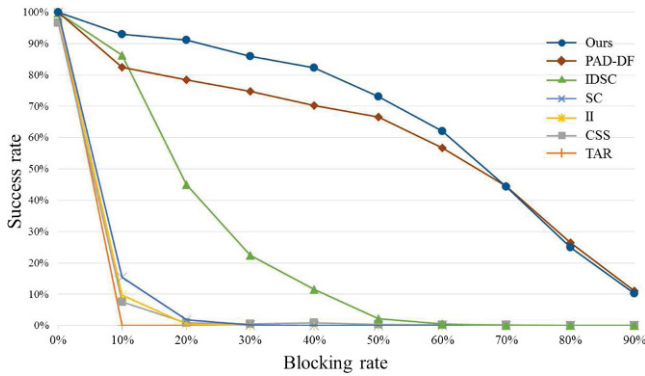


Fig. 11 Matching success versus occlusion.

transformation histogram guarantees that matching pairs with similar transformations are grouped together, while the distance field is greatly affected by the cluttered background.

### 5.3 Tolerance to cluttered edges

We further evaluate robustness to cluttered edges since it is an important factor when detecting shapes in cluttered images. To mimic cluttered edges, we add random arcs to the edge map of a clean shape. We only control the total length of all added arcs with respect to the length of the edges of the original shape. Figure 12 shows a few examples of cluttered instances with different amounts of clutter.

We collected 96 multi-boundary shapes from the Internet. We considered 9 different relative amounts of clutter: 0%, 25%, 50%, 75%, 100%, 150%, 200%, 250%, and 300%. For each shape and clutter level, we created 4 cluttered instances. This gave 3456 cluttered instances in total. Evaluation was performed as in Section 5.2.

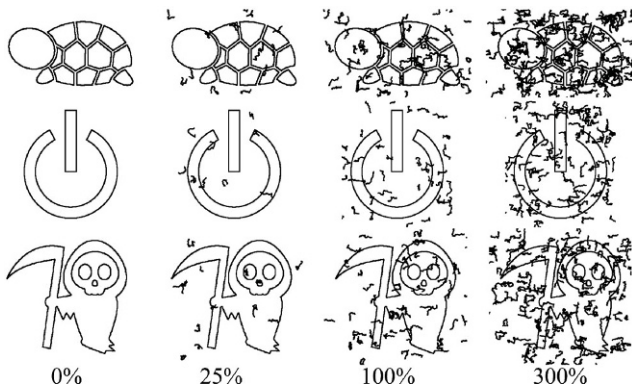


Fig. 12 Cluttered instances, indicating the amount of clutter relative to the original total edge length.

Figure 13 plots the results of our experiment. Since most existing descriptors do not support matching shapes with multiple boundaries and open curves, we only compared our approach with shape contexts and PAD-DF. Our method is much more successful at matching than the other two methods in the presence of cluttered edges. Shape contexts are highly influenced by cluttered edges since they affect the global quantity used for normalization. PAD-DF is also affected by cluttered edges because the distance field is extremely sensitive to noise. Our method works best since cluttered edges are filtered out in the transformation domain.

## 6 Discussion and conclusions

In this paper we have presented a new shape detection approach that robustly detects shapes in cluttered images. By utilizing PAD, our method is able to support shape matching for both open curves and closed shapes, in a scale-invariant, rotational-invariant, and translational-invariant manner. Moreover, our method can detect shape in the presence of partial occlusion.

Our method also has certain limitations. It is sensitive to shape deformation, including change of perspective, because the shape descriptor we use to calculate the transformation does not provide perspective invariance. Figure 14(b) shows a failure in such a case. Although we successfully find the location of the Apple logo in the image, we fail to match the whole shape with a correct transformation. Our method is also sensitive to noise of a kind that leads to changes in curvature of boundary points.

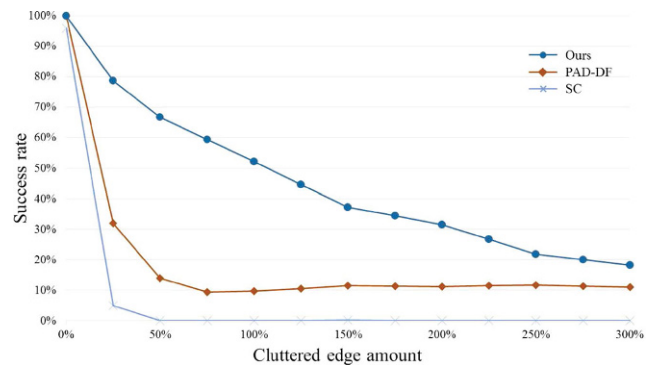
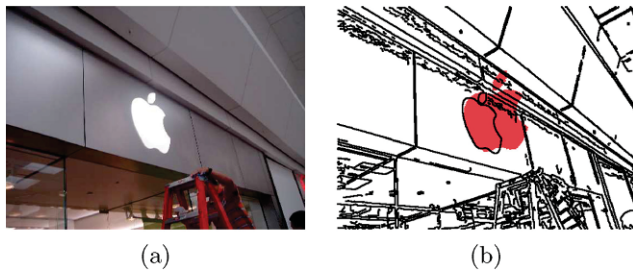


Fig. 13 Matching success versus amount of clutter.



**Fig. 14** Matching under perspective transformation. (a) Input image. (b) Incorrectly detected logo.

## Acknowledgements

This project was supported by the Research Grants Council of the Hong Kong Special Administrative Region, under the RGC General Research Fund (Project No. CUHK 14217516).

## References

- [1] Jacobs, C. E.; Finkelstein, A.; Salesin, D. H. Fast multiresolution image querying. In: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, 277–286, 1995.
- [2] Granlund, G. H. Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers* Vol. C-21, No. 2, 195–201, 1972.
- [3] Persoon, E.; Fu, K. S. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man, and Cybernetics* Vol. 7, No. 3, 170–179, 1977.
- [4] Zhang, D.; Lu, G. Enhanced generic Fourier descriptors for object-based image retrieval. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, IV-3668–IV-3671, 2002.
- [5] Mokhtarian, F.; Mackworth, A. K. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 14, No. 8, 789–805, 1992.
- [6] Mokhtarian, F.; Abbasi, S.; Kittler, J. Efficient and robust retrieval by shape content through curvature scale space. *Image Databases and Multi-Media Search* Vol. 8, 51–58, 1998.
- [7] Alajlan, N.; El Rube, I.; Kamel, M. S.; Freeman, G. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition* Vol. 40, No. 7, 1911–1920, 2007.
- [8] Manay, S.; Cremers, D.; Hong, B.-W.; Yezzi, A. J.; Soatto, S. Integral invariants for shape matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 28, No. 10, 1602–1618, 2006.
- [9] Hong, B. W.; Soatto, S. Shape matching using multiscale integral invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 37, No. 1, 151–160, 2015.
- [10] Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 24, No. 4, 509–522, 2002.
- [11] Ling, H.; Jacobs, D. W. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 29, No. 2, 286–299, 2007.
- [12] Kwan, K. C.; Sinn, L. T.; Han, C.; Wong, T.-T.; Fu, C.-W. Pyramid of arclength descriptor for generating collage of shapes. *ACM Transactions on Graphics* Vol. 35, No. 6, Article No. 229, 2016.
- [13] Chuang, G. C. H.; Kuo, C. C. J. Wavelet descriptor of planar curves: Theory and applications. *IEEE Transactions on Image Processing* Vol. 5, No. 1, 56–70, 1996.
- [14] Tabbone, S.; Wendling, L.; Salmon, J.-P. A new shape descriptor defined on the radon transform. *Computer Vision and Image Understanding* Vol. 102, No. 1, 42–51, 2006.
- [15] Lee, S.-M.; Abbott, A. L.; Clark, N. A.; Araman, P. A. A shape representation for planar curves by shape signature harmonic embedding. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1940–1947, 2006.
- [16] Hu, M.-K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* Vol. 8, No. 2, 179–187, 1962.
- [17] Khotanzad, A.; Hong, Y. H. Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 12, No. 5, 489–497, 1990.
- [18] Belkasim, S. O.; Shridhar, M.; Ahmadi, M. Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition* Vol. 24, No. 12, 1117–1138, 1991.
- [19] Sheng, Y.; Shen, L. Orthogonal Fourier – Mellin moments for invariant pattern recognition. *Journal of the Optical Society of America A* Vol. 11, No. 6, 1748–1757, 1994.
- [20] Bernier, T.; Landry, J.-A. A new method for representing and matching shapes of natural objects. *Pattern Recognition* Vol. 36, No. 8, 1711–1723, 2003.
- [21] Mori, G.; Belongie, S.; Malik, J. Shape contexts enable efficient retrieval of similar shapes. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, I-723–I-730, 2001.

- [22] Mori, G.; Belongie, S.; Malik, J. Efficient shape matching using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 27, No. 11, 1832–1837, 2005.
- [23] Mori, G.; Malik, J. Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, I-134–I-141, 2003.
- [24] Tanase, M.; Veltkamp, R. C. Part-based shape retrieval. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, 543–546, 2005.
- [25] Tănase, M.; Veltkamp, R. C.; Haverkort, H. Multiple polyline to polygon matching. In: *Algorithms and Computation. Lecture Notes in Computer Science, Vol. 3827*. Deng, X.; Du, D. Z. Eds. Springer, Berlin, Heidelberg, 60–70, 2005.
- [26] Pickup, D.; Sun, X.; Rosin, P. L.; Martin, R. R. Skeleton-based canonical forms for non-rigid 3D shape retrieval. *Computational Visual Media* Vol. 2, No. 3, 231–243, 2016.
- [27] Xu, K.; Chen, K.; Fu, H.; Sun, W.-L.; Hu, S.-M. Sketch2Scene: Sketch-based co-retrieval and co-placement of 3D models. *ACM Transactions on Graphics* Vol. 32, No. 4, Article No. 123, 2013.
- [28] Lian, W.; Zhang, L.; Zhang, D. Rotation-invariant nonrigid point set matching in cluttered scenes. *IEEE Transactions on Image Processing* Vol. 21, No. 5, 2786–2797, 2012.
- [29] Thayananthan, A.; Stenger, B.; Torr, P. H. S.; Cipolla, R. Shape context and chamfer matching in cluttered scenes. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, I-127–I-133, 2003.
- [30] Riemenschneider, H.; Donoser, M.; Bischof, H. Using partial edge contour matches for efficient object category localization. In: *Computer Vision – ECCV 2010. Lecture Notes in Computer Science, Vol. 6315*. Daniilidis, K.; Maragos, P.; Paragios, N. Eds. Springer, Berlin, Heidelberg, 29–42, 2010.
- [31] Bai, X.; Li, Q.; Latecki, L. J.; Liu, W.; Tu, Z. Shape band: A deformable object detection approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1335–1342, 2009.
- [32] Cheng, M.-M.; Zhang, F.-L.; Mitra, N. J.; Huang, X.; Hu, S.-M. RepFinder: Finding approximately repeated scene elements for image editing. *ACM Transactions on Graphics* Vol. 29, No. 4, Article No. 83, 2010.
- [33] Toshev, A.; Taskar, B.; Daniilidis, K. Shape-based object detection via boundary structure segmentation. *International Journal of Computer Vision* Vol. 99, No. 2, 123–146, 2012.
- [34] Chi, Y.; Leung, M. K. H. Part-based object retrieval in cluttered environment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 29, No. 5, 890–895, 2007.
- [35] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 39, No. 6, 1137–1149, 2017.
- [36] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788, 2016.
- [37] Gidaris, S.; Komodakis, N. Object detection via a multi-region and semantic segmentation-aware CNN model. In: Proceedings of the IEEE International Conference on Computer Vision, 1134–1142, 2015.
- [38] Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. PAMI-8, No. 6, 679–698, 1986.
- [39] Cui, M.; Femiani, J.; Hu, J.; Wonka, P.; Razdan, A. Curve matching for open 2D curves. *Pattern Recognition Letters* Vol. 30, No. 1, 1–10, 2009.
- [40] Ferrari, V.; Fevrier, L.; Jurie, F.; Schmid, C. Groups of adjacent contour segments for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 30, No. 1, 36–51, 2008.
- [41] Jeannin, S.; Bober, M. Description of core experiments for MPEG-7 motion/shape. MPEG-7, ISO/IEC/JTC1/SC29/WG11/MPEG99 N, 2690, 1999.



**Chu Han** graduated from South China Agricultural University in 2011 with a B.Sc. degree in computer science. He received his M.Phil. degree in computer science from South China University of Technology in 2014, under the supervision of Prof. Xuemiao Xu. He is now pursuing his Ph.D. degree in the Department of Computer Science and Engineering of the Chinese University of Hong Kong, under the supervision of Prof. Tien-Tsin Wong. His current research interests include computer graphics, image processing, pattern recognition, and computer vision.



**Xueting Liu** received her B.Eng. degree from Tsinghua University and Ph.D. degree from the Chinese University of Hong Kong in 2009 and 2014 respectively. She is currently a postdoctoral research fellow in the Department of Computer Science and Engineering of the Chinese University of Hong Kong. Her research interests include computer graphics, computer vision, computational manga and anime, and non-photorealistic rendering.



**Lok Tsun Sinn** graduated from the Chinese University of Hong Kong with a B.Sc. degree in computer science, and is studying for his M.Phil. degree in computer science and engineering in the same department, under the supervision of Prof. Tien-Tsin Wong.



**Tien-Tsin Wong** received his B.Sc., M.Phil., and Ph.D. degrees in computer science from the Chinese University of Hong Kong in 1992, 1994, and 1998, respectively. He is currently a professor in the Department of Computer Science and Engineering of the Chinese University of Hong Kong. His main

research interests include computer graphics, computational manga, precomputed lighting, image-based rendering, GPU techniques, medical visualization, multimedia compression, and computer vision. He received an IEEE Transactions on Multimedia Prize Paper Award in 2005 and a Young Researcher Award in 2004.

**Open Access** The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.