Robust object tracking guided by top-down spectral analysis visual attention

Wanyi Li, Peng Wang, Rui Jiang, Hong Qiao

State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract

Object tracking is an important computer vision task, and existing tracking algorithms have achieved great progress. However, several challenging issues still remain to be solved, such as abrupt motion, longtime complete occlusion and target reappearing after moving completely out of the frame. To address these issues, we propose a top-down spectral analysis visual attention guided tracking method. Given an image sequence, top-down saliency maps for each input image are calculated by introducing top down information, which is learned at the beginning of tracking process, to spectral analysis attention model. Then, guided by the calculated saliency maps, target search is performed by local and global search processes and also with a validation process. Experimental results demonstrate the effectiveness of the proposed method.

1. Introduction

Object tracking is a significant task in many high level image-based applications such as visual surveillance, human–computer interaction and augmented reality. Numerous approaches have been proposed in recent decades [1–6]. However, achieving robust tracking performance still remains a huge challenge. Difficulties in tracking objects can arise due to changes in appearance patterns of both the object and the environment, especially abrupt motion, longtime complete occlusion and target reappearing after moving completely out of the frame [7]. Traditional approaches usually assume that the location of the target would not change significantly from one frame to other, and thus search target object around the location of the target in the last frame. As a result, traditional trackers may fail when abrupt motion or longtime complete occlusion occurs.

Visual attention [8] is one of the key mechanisms of a human visual system which directs processing resources to potentially most relevant visual data, especially directs our gaze rapidly towards objects of interest. As a result, humans can easily achieve robust object tracking. Introducing visual attention mechanism may facilitate realization of stable and humanoid tracking algorithms. Therefore it is very significant to study visual attention based methods for object tracking in complex scenes.

There are two factors guiding human attention [8]: bottom-up attention and top-down attention. Bottom-up attention is purely data-driven and guides the gaze to salient regions in a scene. Regions attracting bottom-up attention are always those with strong contrast or certain uniqueness. Top-down attention is driven by cognitive factors such as pre-knowledge, context, expectations, motivations, and current goals. In the field of computer vision, many computational models have been proposed to simulate humans’ visual attention [9]. Recently, spectral analysis models attracted a lot of interest [10–12]. These approaches derive saliency in the frequency domain in a bottom-up manner and provide state-of-the-art performance in finding salient regions with efficient computation. However, introducing top-down information with considerable importance for tracking to spectral analysis attention models have not been broadly investigated up to now.

In the biological vision literature, it has been suggested that attentional processes are involved in the maintenance of a virtual object during tracking [13] and may facilitate tracking through anticipation or act as an error recovery mechanism [14,15].

For introducing visual attention mechanism into tracking, we firstly propose the following key ideas intuitively:

1. If the motion of object is smooth and complete occlusion does not occur, the target in the current frame will be around the target region of the last frame, and then local search around the last tracking result is enough.
2. When abrupt motion occurs, the moving salient regions attracting more attention during tracking process will be more...
likely to be the target than other regions. Using top-down visual attention computational model, a top-down spatiotemporal saliency map can be generated, and then the target can be obtained by local search or global search based on this saliency map.

3) If the target moves out of the frame or complete occlusion occurs, the tracker should be able to detect and report that the target is lost. Once the target reappears, guided by top-down visual attention, the moving salient regions are more likely to be the target object. In this scenario, the target can be reacquired by global search based on top-down spatiotemporal saliency map.

Due to the proposed ideas and the importance of visual attention in tracking process of human visual system, it is reasonable to introduce top-down visual attention mechanism, especially top-down spectral analysis visual attention, to object tracking. Therefore, this paper presents a novel top-down spectral analysis visual attention guided object tracking method. For simulating top-down visual attention mechanism and calculating saliency maps, we first introduce top-down information to spectral analysis attention model. After calculating saliency maps, local target search and global target search are performed based on these saliency maps. The proposed method can cope with very difficult situations including abrupt motion and target reappearing after long time occlusion or moving completely out of the frame.

The main contributions of this paper are twofold.

- First, we propose a novel robust top-down visual attention guided object tracking framework. This framework includes three components: top-down visual attention, target search and validation. Target search is performed under the guidance of top-down visual attention. The reliability of searched result is validated by validation component. Hence it is robust to abrupt motion and long time target lost caused by complete occlusion or moving out of the frame. Besides, the proposed framework can be extended easily by changing target search strategy and/or target appearance model.

- Second, a top-down saliency detection method which introduces top-down information to spectral analysis attention model is presented to simulate top-down visual attention and calculate saliency maps. To the best of our knowledge, we are the first to propose the top-down spectral analysis visual attention computational model and introduce it in object tracking.

2. The proposed method

The framework of the proposed Visual Attention Guided Tracker (VAGT) is shown in Fig. 1. There are three components in the framework: top-down visual attention, target search and validation. Top-down visual attention calculates spatial saliency maps and spatiotemporal saliency maps. Target search is guided by top-down visual attention and it includes local search and global search performing on calculated saliency maps. Validation component is used to validate the reliability of searched results. A target model and a similarity measure are needed in this component. The target model is represented by color distributions in Hue-Saturation-Value (HSV) color space [13], and the Bhattacharyya coefficient is employed as the similarity measure.

![Visual attention guided tracking framework](image)

*Fig. 1. Visual attention guided tracking framework. In this figure, a light blue rectangle represents a component of the framework and a light yellow rectangle is a subcomponent of a given component.*
Specifically, when a new frame comes, validation component checks whether the target is lost. If not, local search will be performed based on spatial saliency map and spatiotemporal saliency map. Result of local search is validated by validation component. If the target object is lost or result of local search is not reliable, global search is needed and performed based on spatiotemporal saliency map. Fig. 2 summarizes the proposed tracking algorithm according to the framework described in Fig. 1.

2.1. Top-down spectral analysis visual attention

The presented top-down saliency detection method based on spectral analysis attention model is to mimic the top-down visual attention mechanism and calculate saliency maps. A preliminary conference version of the saliency detection method appears in [16]. Four steps are taken in this method: saliency detection using spectral filtering, learning salient features of the target, computing

---

**ALGORITHM 1.** Top-down spectral analysis visual attention guided tracking

**INPUT**
- Image sequence \( f(t) \) with \( t \in \{0, \ldots, T - 1\} \);
- Target region \( R_j \) in \( f(0) \);

**OUTPUT**
- Target state \( S_t \) for each frame \( f(t) \);

**INITIALIZATION**
- \( F_j(f(0)) \leftarrow \) Compute feature maps by Eq. (1)-(8), \( j \in \{R, G, B, Y, I, I_s\} \);
- \( w \leftarrow \) Compute weight vector by Eq. (10);
- \( appearance\_model \leftarrow \) Compute appearance using the way of Sec. 2.3;
- \( S_m\_matched \leftarrow \) True; \( targetLost \leftarrow \) False;
- \( similarity\_thr \leftarrow 0.7 \); \( r \leftarrow 3 \); \( t \leftarrow 1 \);

**WHILE** \( (t < T) \)
- \( F_j(f(t)) \leftarrow \) Compute feature maps by Eq. (1)-(8), \( j \in \{R, G, B, Y, I, I_s\} \);
- \( s\_sm \leftarrow \) Compute spatial saliency map by Eq. (12)-(14);
- \( st\_sm \leftarrow \) Compute spatiotemporal saliency map by Eq. (9) and Eq. (11);
- \( targetLost \leftarrow !S_m\_matched \);
- \( IF \) \( (targetLost) \)
  - \( S_t \leftarrow \) Search target by MeanShift on spatial saliency map \( s\_sm \);
  - \( similarity \leftarrow \) Compute similarity by Eq. (15);
  - \( IF \) \( (similarity >= similarity\_thr) \)
    - \( S_t\_matched \leftarrow \) True; \( CONTINUE \);
- \( ELSE \)
  - \( similarity\_fromspatialmap \leftarrow similarity \);
  - \( S\_fromspatialmap \leftarrow S_t \);
  - \( S_t \leftarrow \) Search target by CamShift on \( st\_sm \);
  - \( similarity \leftarrow \) Compute similarity by Eq. (15);
  - \( IF \) \( (similarity >= similarity\_thr) \)
    - \( S_t\_matched \leftarrow \) True; \( CONTINUE \);
- \( ELSE \)
  - \( result\_local \leftarrow \) Select the better from \( S_t \) and \( S\_fromspatialmap \);
  - \( GO TO \) Global Search;
- \( END \) \( IF \)
- \( ELSE \)
  - \( result\_local \leftarrow \) Null;
  - \( Global Search: \)
    - \( salient\_regions \leftarrow \) Detect salient regions on \( st\_sm \) using the way of Sec. 2.1-(d);
    - \( S_t \leftarrow \) Select best candidate from \( salient\_regions \) and \( result\_local \);
    - \( similarity \leftarrow \) Compute similarity by Eq. (15);
    - \( S_t\_matched \leftarrow \) Evaluate whether \( S_t \) is matched according to \( similarity \).
- \( END \) \( IF \)
- \( t \leftarrow t + 1 \);
- \( END \) \( WHILE \)

---
top-down saliency maps and salient regions detection. We describe more details as follows:

2.1.1. Saliency detection using spectral filtering

The convolution of the image amplitude spectrum with a low-pass Gaussian kernel of an appropriate scale is equivalent to an image saliency detector [12]. The key ideas include (i) spikes in the amplitude spectrum which correspond to repeated patterns; and (ii) repeated patterns that can be suppressed by spectral filtering, and as a result, the saliency pops out from the rest of the image. The procedure of spectral filtering is stated as follows:

To suppress spikes in the amplitude spectrum $|\mathcal{F}\{f(x,y)\}|$ of an image, a Gaussian kernel $g$ is employed in Eq. (1). The saliency map $S(x,y)$ is obtained by reconstructing the 2-D signal using the resulting smoothed amplitude spectrum $A_g$ and the original phase spectrum $P(u,v)$, the reconstruction is computed by inverse Fourier transform shown in Eq. (2).

$$A_g(u,v) = |\mathcal{F}\{f(x,y)\}| \ast g \quad (1)$$

$$S(x,y) = \mathcal{F}^{-1}\{A_g(u,v)e^{i\phi(u,v)}\} \quad (2)$$

In Eqs. (1) and (2), $\mathcal{F}\{f(x,y)\}$ and $\mathcal{F}^{-1}\{f(x,y)\}$ denote Fourier transform and inverse Fourier transform of image $f(x,y)$ respectively, $f \ast g$ is the convolution of $f$ and $g$, and $|x|$ represents the amplitude of complex number $x$.

2.1.2. Learning salient features of the target

Three steps are needed to learn salient features (resulting in a weight vector) for the target to pop out, i.e., (i) calculating feature channels of the image, (ii) calculating the feature saliency map via spectral filtering as stated in previous subsection, and (iii) computing a weight vector for the target object.

(1) Calculating feature channels:

Given an input image $f(t)=f(t)\in\mathbb{R}^T$, where $T$ is the total frame number. $r(t)$, $g(t)$, and $b(t)$ are red, green, and blue channels of $f(t)$. Four broadly tuned color channels [17] are created by Eqs. (3)–(6):

$$R(t) = r(t) - (g(t) + b(t))/2, \quad (3)$$

$$G(t) = g(t) - (r(t) + b(t))/2, \quad (4)$$

$$B(t) = b(t) - (r(t) + g(t))/2, \quad (5)$$

$$Y(t) = (r(t) + g(t))/2 - |r(t) - g(t)|/2 - b(t). \quad (6)$$

The intensity channels and motion channel are calculated by Eqs. (7)–(9):

$$I(t) = (r(t) + g(t) + b(t))/3, \quad (7)$$

$$I_{off}(t) = \max(I(t)) - I(t), \quad (8)$$

$$M(t) = [I(t) - I(t - \tau)], \quad (9)$$

where $\tau$ is the latency coefficient.

(2) Calculating the saliency map of each feature:

Saliency maps of feature channels are calculated by applying spectral filtering to spatial feature channels $R(t)$, $G(t)$, $B(t)$, $Y(t)$, $I(t)$ and $I_{off}(t)$, which are denoted by $F_R$, $F_G$, $F_B$, $F_Y$, $F_I$, $F_{I_{off}}$, $F_t$. These saliency maps of feature channels are called feature maps.

(3) Computing a weight vector for feature maps:

The weight vector $w = (w_R, w_G, w_B, w_Y, w_I, w_{I_{off}})^T$ representing salient features of the target object relate to its surrounding is computed as Eq. (10). The value $w_j$ for feature map $F_j$ is the ratio between the mean saliencies of target region (denoted by $T(0)$) and background (denoted by $f(0), T(0)$):

$$w_j = \frac{\text{mean}(F_j(T(0))))}{\text{mean}(F_j(f(0), T(0))))}; \quad (10)$$

2.1.3. Computing top-down saliency maps

Top-down saliency maps are calculated by introducing top-down information to spectral analysis attention model. The top-down information is included in the previously learned weight vector, and the spectral analysis attention model is reflected by feature maps computed by spectral filtering.

The top-down spatial saliency map $S_{td}$ and the top-down spatiotemporal saliency map $STS_{td}$ are calculated as follows: The computing procedure of $S_{td}$ is similar to VOCUS [18], and $STS_{td}$ is computed as the entrywise product of $S_{td}$ and motion map $M$, see Eq. (11). The motion map can be calculated by Eq. (9).

$$STS_{td} = S_{td} \cdot M, \quad (11)$$

where $A \cdot B$ indicates the entrywise product of $A$ and $B$.

The computing procedures of top-down spatial saliency map $S_{td}$ are stated briefly as follows. When a new frame comes, the previously learned weight vector $w$ is used to weight the feature maps calculated by spectral filtering. Depending on the values $w_j$, the feature maps are employed to generate the excitation map $E$ or the inhibition map $I$. $E$ is the weighted sum of all feature maps $F_j$ that are important for the target region, i.e., $E = \{w_j > 1\} \in \{R, G, B, Y, I, I_{off}\}$:

$$E(f(t)) = \sum_{j \in T_{td}} w_j F_j(f(t)). \quad (12)$$

The inhibition map $I$ considers the features less present in the target region than in the background, i.e., $I = \{w_j < 1\} \in \{R, G, B, Y, I, I_{off}\}$:

$$I(f(t)) = \sum_{j \in T_{td}} (1/w_j) F_j(f(t)). \quad (13)$$

Feature maps with $w_j = 1$ are trivial for the target thus ignored. The top-down spatial saliency map $S_{td}$ results from the difference of $E$ and $I$ and a clipping of negative values:

$$S_{td}(f(t)) = \max(E(f(t)) - I(f(t)), 0). \quad (14)$$

2.1.4. Salient regions detection

In summary, salient regions are detected by iteratively applying CamShift [19] to spatiotemporal saliency map and utilizing the inhibition of return mechanism [17]. The detection procedure is shown in Fig. 3.

Specifically, at any given time, the maximum of the spatiotemporal saliency map defines the most salient image location, this location is taken as initial window location, and target size of last reliable tracking result is taken as initial window size. After obtaining a salient region, the intensity values of all pixels within the current salient region are all set to zero to mimic the inhibition of return mechanism [17]. This iteration will stop when the number of salient regions reaches a predefined value $m_{max}$ or the maximum of the saliency map is less than the predefined threshold value $\epsilon$. These salient regions are denoted by $SR = (SR_1, SR_2, ..., SR_m)$, here $m$ is the number of salient regions. An example for salient region detection is shown in Fig. 4(b).

2.2. Target search

There are two processes in target search, i.e., local search and global search. They are all guided by top-down visual attention
ALGORITHM 2. Salient regions detection combines iterative CamShift and inhibition of return mechanism

INPUT
Spatiotemporal saliency map $\text{STS}_{td}$, size of target object $\text{size}_{\text{target}}$, the maximum number of salient regions $m_{\text{max}}$, predefined threshold value of saliency $\epsilon$.

OUTPUT
Detected salient regions, $SR = \{SR_1, SR_2, \ldots, SR_m\}, m \leq m_{\text{max}}$.

INITIALIZE
$i = 1$, $SR = \emptyset$

WHILE ($i \leq m_{\text{max}}$)

($\text{maxValue}, \text{maxLoc}$) $\leftarrow$ search the maximum of the current spatiotemporal saliency map $\text{STS}_{td}$ and the corresponding location;

IF ($\text{maxValue} \geq \epsilon$)

$\text{windowLocation} \leftarrow \text{maxLoc}$

$\text{windowSize} \leftarrow \text{size}_{\text{target}}$

$sr_{\text{current}} \leftarrow \text{CamShift}(\text{STS}_{td}, \text{window})$

$SR \leftarrow SR \cup \{sr_{\text{current}}\}$

$\text{STS}_{td}(sr_{\text{current}}) \leftarrow 0$

ELSE

RETURN;

END IF

$i = i + 1$;

END WHILE

Fig. 3. The salient regions detection procedure.

Fig. 4. Target search guided by top-down visual attention. (a) Target object (top right) and searched result (red rectangle); (b) Top-down spatial saliency map and detected salient regions (green rectangles).
mechanism. Local search is realized by applying MeanShift [20] on spatial saliency map and CamShift [19] on spatiotemporal saliency map. Global search is done by selecting the best candidate region from detected salient regions and result of local search. A simple example for target search guided by top-down visual attention is shown in Fig. 4.

2.3. Validation

Validation component is used to determine whether the target is lost and whether a search result is reliable. A target model and a similarity measure are needed in this component. Target model is represented as color distributions in Hue-Saturation-Value (HSV) color space [21]. And the complete histogram is composed of \( N = N_h N_s + N_v \) bins. \( N_h N_s \) bins are for HS histogram, and use only the pixels with saturation and value larger than two thresholds which are respectively set as 0.1 and 0.2 in the experiments. The reason for using thresholds is that color information is only reliable when both the saturation and the value are not too small. Another \( N_v \) bins are for the remaining “color-free” pixels which retain a crucial information when tracked regions are mainly black and white. We employ the Bhattacharyya coefficient as the measure of the similarity between two color distributions: \( p = \{ p^{(1)}, \ldots, p^{(m)} \} \) and \( q = \{ q^{(1)}, \ldots, q^{(m)} \} \). It is defined as

\[
\rho(p, q) = \sum_{u=1}^{m} \sqrt{p^{(u)} q^{(u)}}
\] (15)

In Eq. (15), the larger \( \rho \) is, the more similar the distributions are. When two distributions are identical, we get \( \rho = 1 \), which indicates a perfect match.

3. Experiments and results

Our tracker is implemented in C++ which runs at 5 fps on a PC with an Intel 2.93 GHz Dual Core CPU and 2 GB memory. For each sequence, the target region \( R_0 \) in \( f(0) \) is initialized as the ground truth position in the first frame. Empirically, the latency coefficient \( \tau \) in Eq. (9) is set to 3, the maximum number of salient regions \( m_{max} \) and the saliency threshold \( \varepsilon \) in Fig. 3 are set to 5 and 0.1 respectively.

To evaluate our approach (VAGT), we compare it with original MeanShift tracker [20] and color-based Particle Filter tracker [21] on several challenging sequences.

3.1. Qualitative comparison

Figs. 5–7 show the comparative tracking results of all methods on several challenging videos. Fig. 5 shows the comparison performances when dealing with abrupt motion. The color-based Particle Filter tracker and original MeanShift tracker drift away from the target since the 83rd frame, while the proposed tracker can always lock onto the object. Fig. 6 shows the comparison results when the target is under longtime occlusion. After losing the target at the 53rd frame with completed occlusion, the proposed tracker recovers to capture the object while the others fail. Fig. 7 shows tracking result of “lemming” sequence with abrupt motion and complete occlusion. While proposed approach succeeds during the tracking period, the color-based particle filter tracker is distracted by a similar image region at the 600th frame and the original MeanShift tracker loses the target at the 300th frame. In these sequences, the proposed method VAGT can always lock onto the object while the color-based Particle Filter tracker and original MeanShift tracker lose target to some extent.

Fig. 5. Comparison of tracking results under abrupt motion between different trackers. The frames 2, 83, 410, and 412 are shown. (a) Results of proposed tracker. (b) Results of color-based particle filter tracker. (c) Results of original MeanShift tracker.
Fig. 6. Comparison of tracking results under longtime occlusion between different trackers. The frames 2, 50, 53, and 80 are shown. (a) Results of proposed tracker. (b) Results of color-based particle filter tracker. (c) Results of original MeanShift tracker.

Fig. 7. Comparison of tracking results for “lemming” sequence. The frames 2, 300, 600, and 900 are shown. (a) Results of proposed tracker. (b) Results of color-based particle filter tracker. (c) Results of original MeanShift tracker.
3.2. Quantitative evaluation

For a quantitative evaluation, the spatial overlap metric [22] is used. Given tracked bounding box $S_t$ and ground truth bounding box $S^T_t$, the overlap score is defined as $SC = |S_t \cap S^T_t|/|S_t \cup S^T_t|$, where $\cap$ and $\cup$ represent the intersection and union of two regions respectively, and $|\cdot|$ denotes the number of pixels in the region. The target in frame $t$ is accurately estimated if $SC_t \geq SC_{thr}$ where the threshold $SC_{thr}$ is set to 0.3 in our experiments.

Table 1 shows the tracking performances scores. The accuracy is defined as the ratio between the number of frames on which object location is accurately estimated and the total number of frames on which the object is not completely occluded in the sequence. Our visual attention guided tracker outperforms other two trackers for all the three sequences.

For time complexity, all methods are performed on a PC with an Intel 2.93 GHz Dual Core CPU and 2 GB memory, our tracker is implemented in C++ and runs at about 5 fps on average without code optimization. It is hopeful to run in real time after a properly optimizing.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Total number of frames except those with complete occlusion</th>
<th>The number of frames on which the object is estimated accurately</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrupt motion sequence</td>
<td>643</td>
<td>76</td>
<td>11.8%</td>
</tr>
<tr>
<td>Longtime occlusion sequence</td>
<td>100</td>
<td>51</td>
<td>51%</td>
</tr>
<tr>
<td>“Lemming” sequence</td>
<td>1319</td>
<td>404</td>
<td>30.6%</td>
</tr>
</tbody>
</table>

Notes: MS=Mean-shift, C_PF=Color-based Particle Filter, VAGT=Visual Attention Guided tracker.

4. Conclusions

We propose a novel top-down spectral analysis visual attention guided object tracking method. Experimental results show the following key points. Firstly, the presented method is robust to abrupt motion and longtime occlusion even completely out of the frame. Secondly, the proposed tracker outperforms the traditional MeanShift and color-based Particle Filter trackers with the same appearance model on several challenging video sequences, and this indicates that the robustness comes from guidance of top-down visual attention instead of the appearance model.

Based on the proposed tracking framework, some future work can be extended, such as new search strategies, adaptive appearance models and extensive experiments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under the Grant 61401463, 61100098, 61210009 and 61379097.

References


Wanyi Li received his M.Eng. degree in computer science from the Guizhou University, Guiyang, China, in 2010, and his PhD. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2014. He is currently an Assistant Professor with the Institute of Automation, Chinese Academy of Sciences. His current research interests include computer vision and machine learning.
Peng Wang received his B.Eng. degree in electrical engineering and automation from the Harbin Engineering University, Harbin, China, in 2004, his M.Eng. degree in automation science and engineering from the Harbin Institute of Technology, Harbin, in 2007, and his Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010. He is currently an Associate Professor with the Institute of Automation, Chinese Academy of Sciences. He has worked on computer vision, image processing, and intelligent robots. His current research interests include computer vision, human–robot interaction and robotics.

Rui Jiang received her B.Sc. degree in Computational Mathematics from the Xi’an Univ. of Technology (XUT), Xi’an, China, in 2006 and her M.Sc. degree in Applied Mathematics from the Xi’an Jiaotong Univ. (XJTU), Xi’an, China, in 2010. She is now a Ph.D. Candidate at the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China. Her current research interests include machine learning, pattern recognition and computer vision.

Hong Qiao received her B.Eng. degree in hydraulics and control and her M.Eng. degree in robotics from the Xi’an Jiaotong University, Xi’an, China, her M.Phil. degree in robotics control from the Industrial Control Center, University of Strathclyde, Strathclyde, UK, and her Ph.D. degree in robotics and artificial intelligence from De Montfort University, Leicester, UK in 1995. She was a University Research Fellow with De Montfort University from 1995 to 1997. She was a Research Assistant Professor from 1997 to 2000 and an Assistant Professor from 2000 to 2002 with the department of manufacturing engineering and engineering management, City University of Hong Kong. Kowloon, Hong Kong. Since January 2002, she has been a Lecturer with the School of Informatics, University of Manchester, Manchester, UK. Currently, she is also a professor with the Laboratory of Complex Systems and Intelligent Science, Institute of Automation, Chinese Academy of Sciences, Beijing, China. She first proposed the concept of “the attractive region in strategy investigation,” which has successfully been applied by herself in robot assembly, robot grasping, and part recognition. The work has been reported in Advanced Manufacturing Alert (Wiley, 1999). Her current research interests include information-based strategy investigation, robotics and intelligent agents, animation, machine learning (neural networks and support vector machines), and pattern recognition.

Dr. Qiao is also a member of the Program Committee of the IEEE International Conference on Robotics and Automation from 2001 to 2004. She is currently a member of Administrative Committee of the IEEE Robotics and Automation Society, also the Associate Editors of the IEEE Transaction on Systems, Man, and Cybernetics –Part B and the International Journal of Social Robots.