Comfortable Subjective Duration and User Experience of Face Recognition

Tingting Gan(✉) and Chengqiang Yi

Baidu, Beijing, People's Republic of China
{gantingting,yichengqiang}@baidu.com

Abstract. Face recognition, as an important biometric technique for personal identification, has been widely used in many departments as government, public security, banking, securities, taxation and army. However, most previous research paid more attention on technology in accuracy and speed and ignored user experience, which was our focus. We evaluated user experience of our and competing products, furthermore, quantitatively analyzed comfortable subjective duration of three stages called face detection, blink detection and picture-taking, adopting tolerance experiment and usability test. The result revealed that comfortable subjective duration of three stages were 1–2.5 s, 0.8–1.8 s, 0–0.7 s. Combined with the result of usability test, we optimized UE/UI design to enhance the user experience.

Keywords: Face recognition · Comfortable subjective duration · User experience

1 Introduction

Face recognition was the process of determining the size, location, position and orientation of human face from the input image, the research of which can be traced back to the 1960s. And it has become increasingly mature after decades of tortuous development [1, 2]. Face recognition was an important biometric technique for personal identification, has been widely used in many departments as government, public security, banking, securities, taxation and army. Most previous research focused more on technology in accuracy and speed [3–6], such as Galton proposed the use of the key points and their distances of the human face to represent the feature vector of the human face.

However, many researchers revealed that objective stimuli and subjective experience tended to be always inconsistent [7, 8]. Less studies assessed the user subjective evaluation in the process of face recognition in order to achieve not only the goal of a fast and accurate identification but only a good user experience. On the basis of this problem, our research paid more attention on the following questions:

1. What was the objective performance of our and competing product in terms of pass rate and recognition length after contrasting test?
2. What was the user's subjective evaluation of our and competing product? And was the subjective evaluation consistent with objective performance? If there is inconsistency, what was it and we needed to find out the reason in order to optimize UE/UI design.

A. Marcus and W. Wang (Eds.): DUXU 2017, Part III, LNCS 10290, pp. 238–248, 2017.
DOI: 10.1007/978-3-319-58640-3_17

3. In the face recognition, what was the most appropriate subjective duration from the user perspective?

The current purpose might provide a direction of the further optimization direction on the users' subjective experience of face recognition through a comparison test and a rigorous experiment.

2 Research Method

2.1 Procedure

The whole study was divided into two parts. The first part was similar to usability test to obtain the objective performance of our and competing product. The second part was related to tolerance experiment, which was widely used in psychological experiment and aimed to obtain the intolerable length of time [9, 10].

Specifically, in the usability test we required our participants to conduct face recognition in our and competing products in sequential order. To balance order effect, half of the participants did our product first and another half did competing product first. Researchers recorded the pass rate and recognition length. And we also collected feedback of participants on the subjective experience during use.

In the tolerance experiment, participants were required to give an assessment on a five point Likert scale, namely 1-very unacceptable, 2-a little unacceptable, 3-moderate, 4-a little acceptable, 5-very acceptable, to a given duration of face recognition and then we calculated the duration interval of extremely well (4 points or more), moderately well (3–4 points), just passable (2–3 points) and completely intolerable (2 points or less). In order to facilitate targeted optimization, we divided the whole face recognition into three stages. The first stage was face detection, from the action of clicking the start button to the appearance on the page saying "face detected". The second stage was blink detection, from the appearance saying "please blink" to the appearance saying "blink detected". The third stage was taking a photo, from the appearance saying "please be ready to take photos" to the appearance saying "successfully completed" (see Fig. 1).

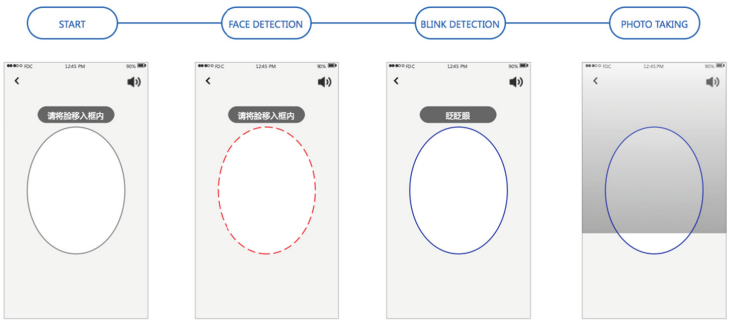

Fig. 1. The three stages of face recognition

We invited programmers to develop different demos in terms of duration for three stages respectively. As the first stage for example, it had 24 demos from 0.5 s to 15 s. The participant was given a demo at a time, the duration of which was programmed in advance, unrelated with actual behavior of the participant without his knowing. And he was required to evaluate his subjective acceptability for this given demo in a Likert scale. Every participant should evaluate all demos in an ascending or descending order. Similarly, half of them followed the former and the other half adhered to the latter to avoid sequential effect.

2.2 Material

Due to no need of develop demos for the first usability test, we adopted online version of our and competitive product to obtain comparable result in a real-world environment. However, the second part of tolerance experiment adopted demos which were scheduled time in advance. Before this, it was unknown what the appropriate quantity of demos and duration gap between the two adjacent demos were, which was determined in a pre-experiment. It was found out that 80% of participants cannot be tolerated if the duration exceeded 11 s, 9 s and 5 s for the three stages, namely face detection, blink detection and picture taken. As a precautionary measure, the duration for the three stages increased to 15 s, 11 s, 7 s. Duration gap between the two adjacent demos must fit in with the principle of neither so long to miss key turning point nor so short to unable to perceive discrepancy. Combining the previous studies which found out difference threshold distributed from 0.65 s to 1 s, we choose 0.5 s for the duration gap between the two adjacent demos.

Finally, the number of demos for first stage, face detection, from 0.5 s to 15 s, was 24. The number of demos for second stage, blink detection, from 0.5 s to 11 s, was 22. The number of demos for third stage, picture taken, from 0.5 s to 7 s, was 14. All demos in the same stage were equal excepted for duration (see Table 1).

2.3 Participants

We invited 10 participants for the pre-experiments and 28 for the formal experiment and usability test, whose occupations involved student, self-employed person and employee. Their age distributed from 18 to 25, with the mean value of 21. To avoid an impact of mobile phone models, we selected 16 IOS users and 12 android users.

2.4 Results

Objective Performance. The accurate recognition rate of both our and competing products achieved to 90%. Our product had the advantage over the other in the whole objectively time consuming (see Table 2).

However, the result that the subjective evaluation from users was worse than competing product was entirely unexpected. Further analysis revealed the most two

Table 1. All demos for three stages.

First stage (24 demos)	Second stage (22 demos)	Third stage 14 demos)
0.5 s	0.5 s	0.5 s
1.0 s	1.0 s	1.0 s
1.5 s	1.5 s	1.5 s
2.0 s	2.0 s	2.0 s
2.5 s	2.5 s	2.5 s
3.0 s	3.0 s	3.0 s
3.5 s	3.5 s	3.5 s
4.0 s	4.0 s	4.0 s
4.5 s	4.5 s	4.5 s
5.0 s	5.0 s	5.0 s
5.5 s	5.5 s	5.5 s
6.0 s	6.0 s	6.0 s
6.5 s	6.5 s	6.5 s
7.0 s	7.0 s	7.0 s
7.5 s	7.5 s	
8.0 s	8.0 s	
8.5 s	8.5 s	
9.0 s	9.0 s	
9.5 s	9.5 s	
10.0 s	10.0 s	
10.5 s	10.5 s	
11.0 s	11.0 s	
13.0 s		
15.0 s		

Table 2. The whole objective time consuming of our and competing products

Unit: s	iOS	Android
Our product	9.57	9.75
Competing product	10.82	12.05

important reasons were unsatisfactory interactive details and subjective perceived duration.

The former consisted of the following four details (see Table 3).

① The prompt frame of face detection included too many outlines of face and shoulder for user to manipulate in the given frame.
② Lack of clear and vivid hint, users cannot go forward according to our design.
③ When confronted the failure of face detection, users were at a loss what to do due to the inappropriate time and content of the prompt.
④ Our product was in short of interest, exquisite feeling and specialty.

Table 3. Unsatisfactory interactive details of our product

	Our weaknesses	Competing strength
Frame for prompt	1. The prompt frame of face detection included too many outlines of face and shoulder for user to manipulate in the given frame 2. The frame was so small that users needed to adjust mobile phone in a far position which was inappropriate for self-portrait	1. The prompt frame included only face 2. The frame was big enough for self-portrait
Prompts	1. The visual focus was in the prompt for frame, but word hit for prompt was on the top of frame. So it was difficult for users to observe. 2. Lack of Voice prompt 3. Visual cues were not obvious.	1. Obvious voice and visual prompt
Prompt of failure	1. When countering the failure of detection, users received the prompt too late 2. The contents was in common use without basing on the situation	1. The prompt appeared timely 2. The contents of the prompt was to a point 3. When monitoring a poor-light environment, the screen of competing product can be adjusted automatically
Visual effect		1. Users perceived less time with more dynamic effects 2. Full of exquisite feeling and specialty

The latter illustrated the inconsistence between subjective comfortable duration and objective duration. So what was the most subjective comfortable time of face cognition?

2.5 The Standard of Comfortable Duration

Every participant evaluated all demos of different duration and so we can calculate the mean value of all participants at every duration. Furthermore, the percentage of the participants who felt extremely well, moderately well, just passable at every duration can be calculated respectively.

In consideration of the two results, we can come to an agreement of a standard duration.

The result revealed that users felt extremely well from 1 s to 2.5 s (see Fig. 2), in the interval of which 70% of participants perceived comfortable in the face detection stage (see Fig. 3).

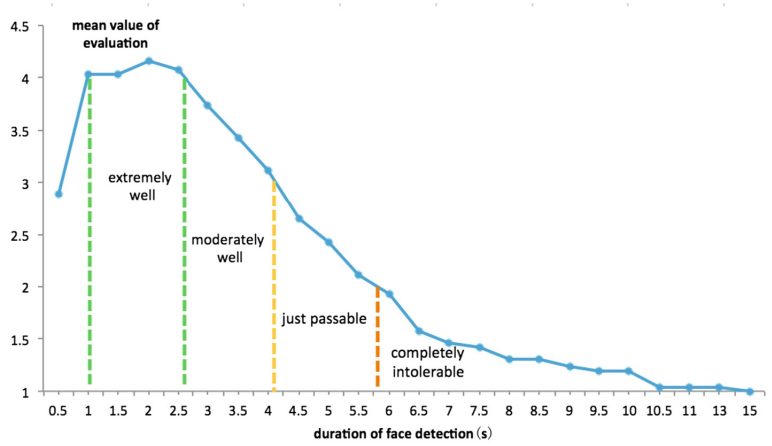

Fig. 2. The interval of users felt extremely well in the face detection stage 1-very unacceptable, 2-a little unacceptable, 3-moderate, 4-a little acceptable, 5-very acceptable

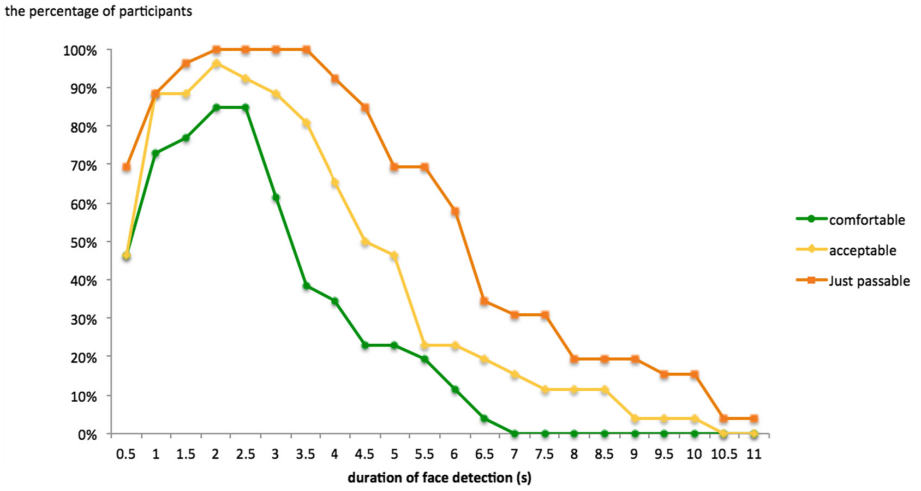

Fig. 3. The percentage of users who felt comfortable, acceptable and just passable in the face detection stage

Similarly, users felt extremely well from 0.8 s to 1 s (see Fig. 4), in the interval of which 80% of participants perceived comfortable in the blink detection stage (see Fig. 5).

And in the picture-taking stage the result was within 0.7 s (see Fig. 6). and 80% (see Fig. 7).

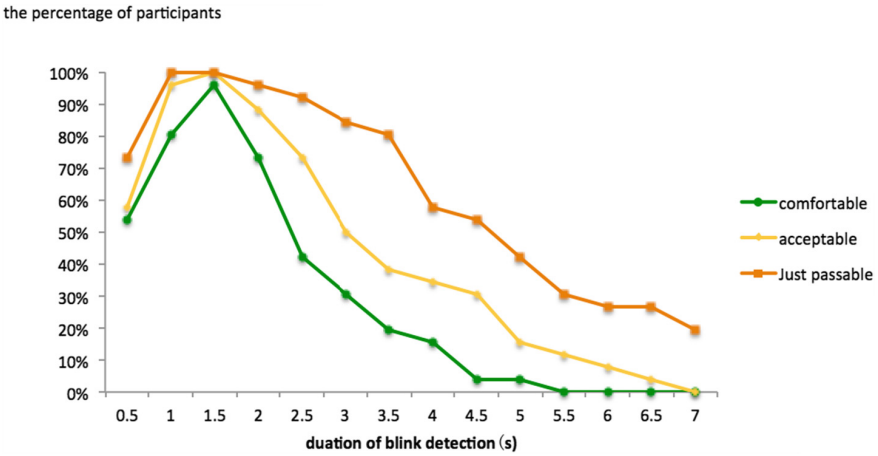

Fig. 4. The interval of users felt extremely well in the blink detection stage

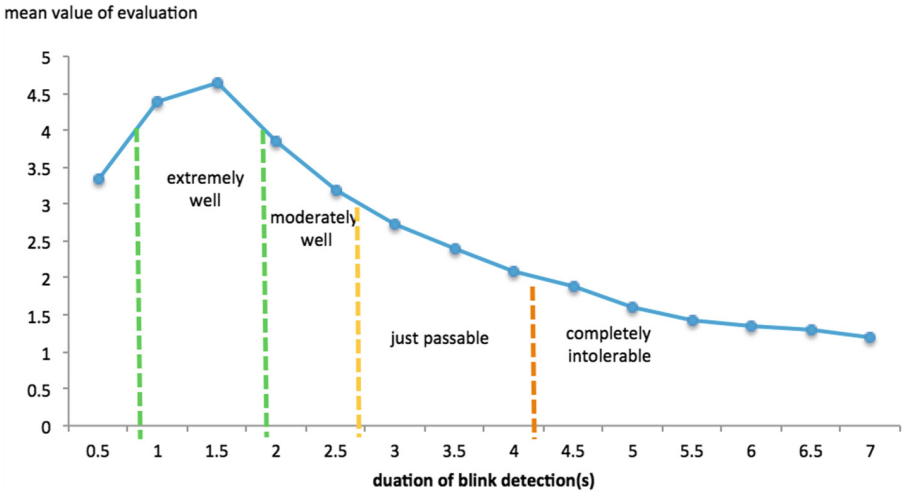

Fig. 5. The percentage of users who felt comfortable, acceptable and just passable in the blink detection stage

3 Discussion

In the interaction design of user interface, designers needed to guide users to go forward in line with your expectation to achieve product goal through the way of making the focal key points stand out. However, there existed some problems in the initial design.

We expected that users interacted in the sequence of hits-frame-other, however, people focused all their attention on the frame (see Fig. 8).

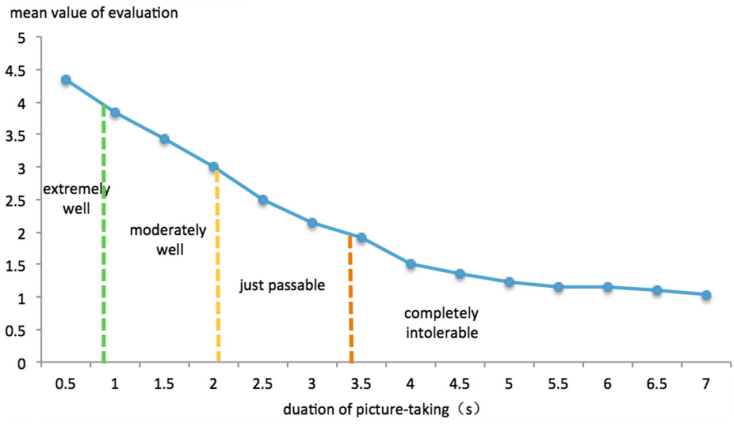

Fig. 6. The interval of users felt extremely well in the picture-taking stage

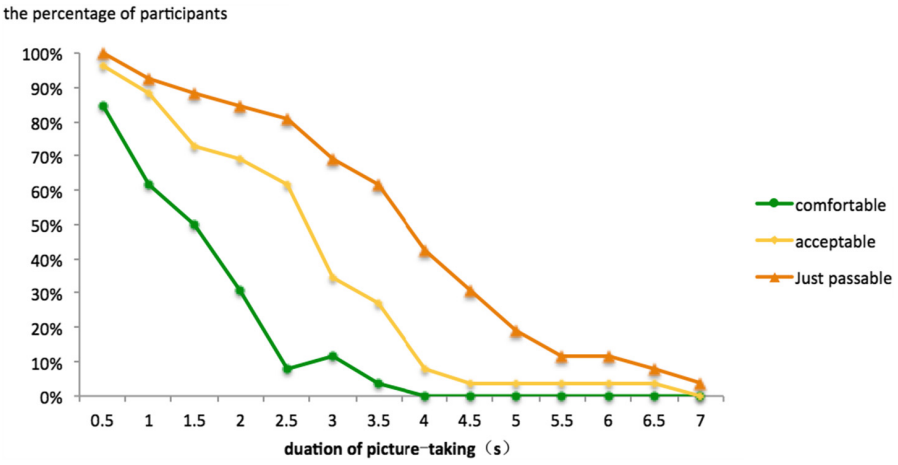

Fig. 7. The percentage of users who felt comfortable, acceptable and just passable in the picture-taking stage

So how to make the focal key points stand out was an important question. As our usual practice went, contents, color, size and distance might be in use (see Fig. 9). However, face recognition, as a novel Interactive mode, was difficult for users to understand. So we adopted more visual motion and sound in a subsequent improvement (see Fig. 10).

And we found out that in the face and blink stage, the subjective experience of users and the objective duration were inconsistent. Our data (see Fig. 10) revealed that current status of our product cannot meet users' demand in terms of degree of comfort. So we adjusted the time length of visual dynamic effect to increase the duration of detection by 1 s to push forward more participants' subjective experience into the comfortable interval (see Figs. 11 and 12).

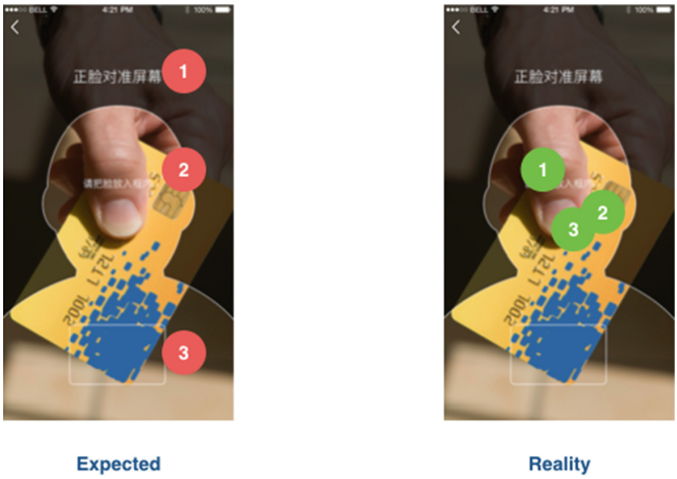

Fig. 8. The expected sequence vs. the real sequence

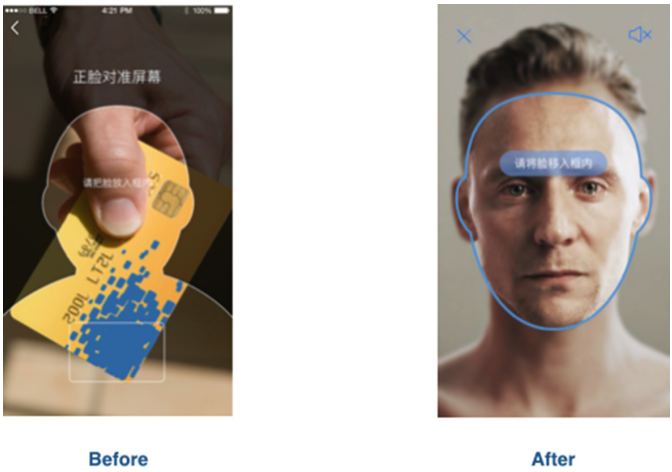

Fig. 9. The product of before vs. after (Color figure online)

In the third stage, the longer the duration was, the better users perceived. So the performance of the recognition should be promoted in respect of technology.

As a whole, the present study tried to build a standard duration of face recognition in terms of users' subjective experience instead of objective technology, using usability test and tolerance experiment. We hope our result can give a reference to future optimization of related product.

However, we conducted the experiment in a normal environment, without considering circumstance differences, such as did the credit and financing circumstance do some different effect on the perceived duration? So we will conceive a more well-designed experiment to explore the subjective duration in more situations and more group division.

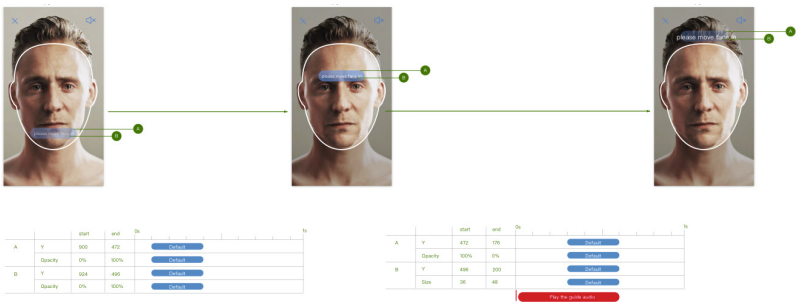

Fig. 10. Visual motion and sound in the face recognition

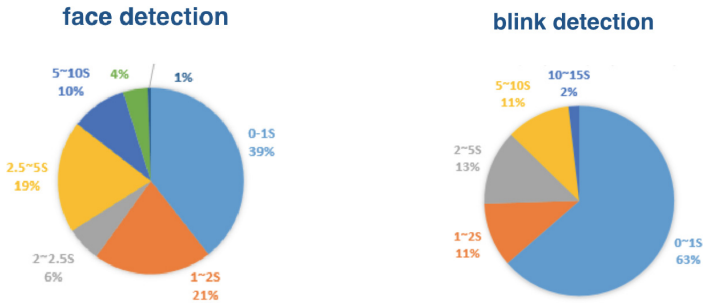

Fig. 11. In the face detection and blink detection before optimization

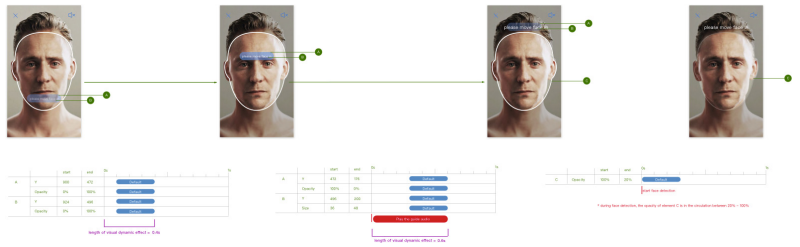

Fig. 12. Adjusting the time length of visual dynamic effect to increase the duration of detection by 1 s

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces, a survey. Proc. IEEE **83**(5), 705–740 (1995)
2. Zhou, J., Lu, C.-Y., Zhang, C.-S., et al.: A survey of automatic face recognition. Acta Electronica Sinica **28**(4), 102–106 (2000)
3. Galton, F.: Numeralised profiles for classification and recognition. Nature **83**(2109), 127–130 (1910)

4. Bledsoe, W.W., Chan, H.: A man-machine facial recognition system: some preliminary results. Technical report, PRI 19A, Panoramic Research Incorporated, Palo Alto, USA (1965)
5. Wang, Z.H., Yuan, H., Jiang, W.T.: A face recognition algorithm based on composite gradient vector. Acta Automatica Sinica **37**(12), 1445–1454 (2011). (in Chinese)
6. Sharma, A., Haj, M.A., Choi, J., et al.: Robust pose invariant face recognition using coupled latent space discriminant analysis. Comput. Vis. Image Underst. **116**(11), 1095–1110 (2012)
7. Posner, M.I.: Orienting of attention. Q. J. Exp. Psychol. **32**, 2–25 (1980)
8. Eriksen, C.W., Yeh, Y.Y.: Allocation of attention in the visual field. J. Exp. Psychol.: Hum. Percept. Perform. **11**, 583–597 (1985)
9. Martinez-Trujillo, J.C., Treue, S.: Feature—based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. **14**, 744–751 (2004)
10. Duncan, J.: Selective attention and the organization of visual information. J. Exp. Psychol.: Gener. **113**, 501–517 (1984)