The Effect of Object Features on Multiple Object Tracking and Identification

Tianwei Liu[1,2], Wenfeng Chen[1,*], Yuming Xuan[1], and Xiaolan Fu[1]

[1] State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
{chenwf,xuanym,fuxl}@psych.ac.cn
[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China
liutian_wei@163.com

Abstract. Tracking of multiple objects is challenging for computer vision system under certain circumstance. We investigated this problem with human observers. In our experiment, observers were asked to track multiple moving items as well as to maintain their identities. We found that the capacity of maintaining multiple moving object identities of human is about three to four items, and uniqueness improves the general tracking and ID performance. It also showed that observers' capacity of ID task was dependent on feature type, and suggests that a less resource-demanding process of identity-related feature would lead to more effective improvement on tracking. These results provide some indications for the design of computer vision system which involves human monitoring, and suggest that creating a featural space to map the identity of multiple objects may aid the automatic object tracking.

1 Introduction

It is important to identify and maintain information of multiple objects that move through an environment. The Multiple Object Tracking (MOT) paradigm is popularly used to study how our visual system tracks multiple moving objects [1].

The research has generated anti-intuitive findings concerning the role of distinct feature on tracking. Objects are commonly distinguished by features. Yet relative to object locations, object features such as color and shape are often lost easily during tracking [2]. Intuitively, individuating objects should facilitate tracking. Horowitz et al [3] addressed similar question that whether unique objects help or hinder MOT performance in comparison with identical objects (animals) and found the unique object advantage. However, this unique advantage was not always observed [4]. Furthermore, it remains intact for the role of distinct feature on identification based on identity-location binding during tracking. So in the present work, we tested that whether distinct feature information could improve tracking and identification.

* Corresponding author: chenwf@psych.ac.cn

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2009, LNAI 5639, pp. 206–212, 2009.

2 Methods

2.1 Participants

Twenty-four participants (12 female) aged 19-25 (Mean 21.1) participated in the experiment as paid volunteers. All participants had normal or corrected to normal vision.

2.2 Stimuli

The stimuli were presented on a 17-inch color CRT monitor controlled by E-prime 1.1. The refresh rate of monitor was 75 Hz at a resolution of 800×600 pixels. The viewing distance was about 60 cm.

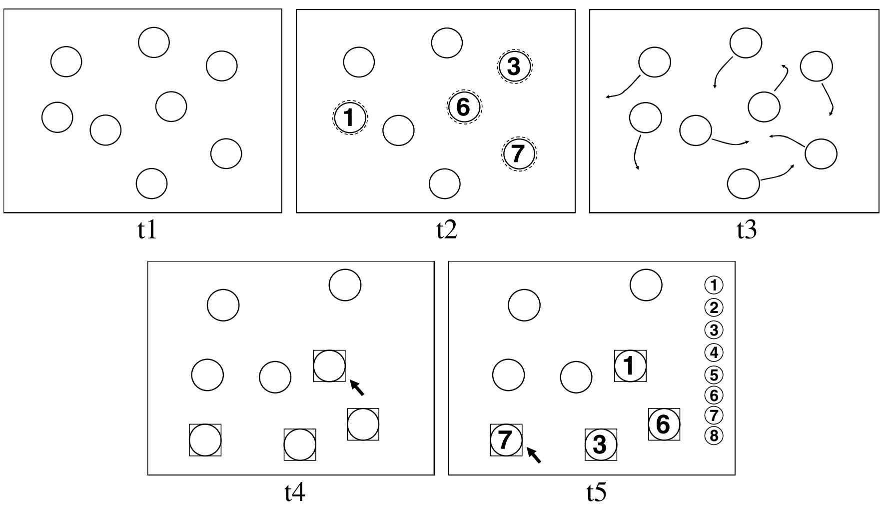

Fig. 1. Sequence of events in a trial in the identical and number condition. At the beginning of a trial, eight disks were shown at their random initial positions (t1). Then four of them flashed on and off to identify they were targets meanwhile inside each of these four disks a number randomly selected from 1 to 8 without replacement would be displayed. (t2). The numerical labels on the targets disappear and all items began to move for 5 or 10 seconds (t3). The disks stopped moving and observers selected out the 5 targets (t4). Observers assigned a number to each target (t5).

The objects of tracking were eight identical grey disks and four of them were targets. Each disk was 47 pixels in diameter and was surrounded by a white border of 2 pixels. The screen background was black and the movement was restricted in the center area subtending about 18.2 × 18.2° of visual angle (moving area). The initial positions of the eight disks were random selected on the screen with the constraint that each disk had to be at least 1° from the edges of the moving area and at least 1.1° from each other. The velocity of the items varied between 1.9 and 2.8° / s with a mean of 2.4 ° / s.

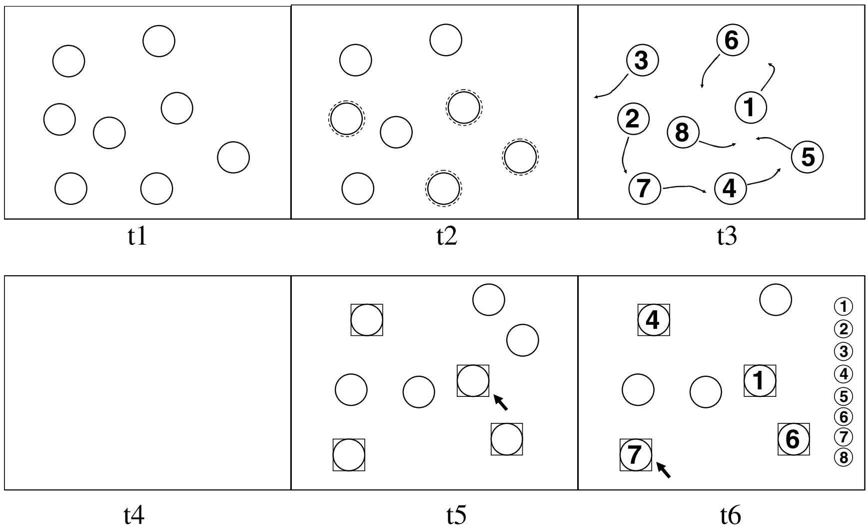

Fig. 2. Sequence of events in a trial in the unique number condition. At the beginning of a trial, eight disks were shown at their random initial positions (t1). Then four of them flashed on and off to identify they were targets (t2). All items began to move for 5 or 10 seconds with a unique number appearing in their centers throughout the movement (t3). A white screen flashed for 100ms to eliminate the visual afterimage (t4). Observers selected out the four targets (t5). Observers assigned a number to each target (t6).

2.3 Design and Procedure

There was a mixed design with object differentiability (unique, identical), moving duration (5s, 10s) as within-participant variables and feature type (numeral label, color label) as between-participant variable. Twelve participants were tested in the number condition and twelve in the color condition.

We introduced a variant of the classical MOT procedure. The main change was that an identification task based on identity-location binding was added after tracking task. The essential change for unique condition was that all the items were visually unique during movement because of the numerical or color labels which appeared when they began to move and remained visible until they stopped moving.

At the start of each trial, eight disks appeared on the screen and four of them flashed for 1500 ms designated as targets. Then all items began to move for 5 or 10 seconds. During the moving of the objects, overlap between any items randomly appeared. When disks stopped moving, the participants were instructed to pick out all the items that had flashed before by clicking on them with the mouse (tracking task). Then the list of digits from 1 to 8 (or colors) was displayed nearby the right edge of the screen. Participants were asked to label each of the targets to the correct number that had ever showed at the beginning of the trial (identification task). They could click a target and then click one of the eight numbers. The trial ended after all of targets were labeled. The observer then pressed the spacebar to initiate the next trial. The sequence of events in each trial was shown in Figure 1 and 2.

For unique condition, a number (or color) label randomly selected from 1 to 8 (or blue, coffee, cyan, green, orange, purple, red, yellow color) would be displayed inside each disk during moving. For identical condition, a number (or color) label would be displayed inside each disk when flashing and labels disappeared when moving.

3 Results

Raw accuracy data were transformed into estimated capacities according to high-threshold guessing models devised for each condition. The estimated capacities of tracking and identification were computed according to the formulas of Horowitz et al. [3].

The tracking capacities were shown in Figure 3. A three-way ANOVA revealed that there were significant main effects of object differentiability [$F(1, 22) = 42.44$, $p < .001$] and moving duration [$F(1, 22) = 25.69$, $p < .001$], and interaction between these two variables [$F(1, 22) = 18.92$, $p < .001$], but not significant main effect of feature type and other interactions[$F(1, 22) < 1.20$, $p > .12$]. Simple main effect analysis showed that the difference between moving durations was significant for identical condition [$p < .001$], but was not significant for unique condition [$p = .759$].

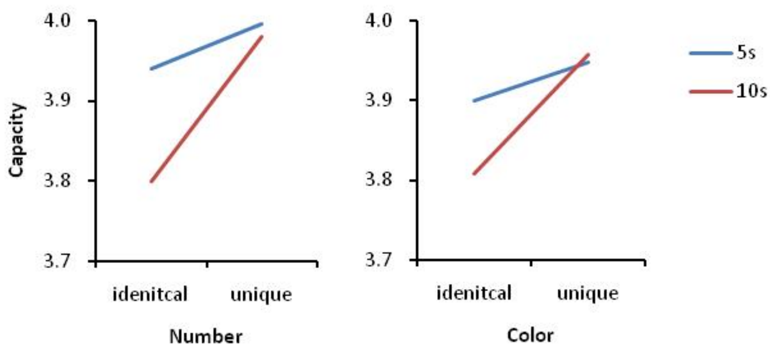

Fig. 3. Tracking capacity as a function of feature type, object differentiability and moving duration

The identification capacities were shown in Figure 4. A three-way ANOVA revealed that there were significant main effects of feature type [$F(1, 22) = 11.11$, $p < .01$], object differentiability [$F(1, 22) = 60.71$, $p < .001$] and moving duration [$F(1, 22) = 19.82$, $p < .001$], and interaction between these latter two variables [$F(1, 22) = 47.85$, $p < .001$], but not significant main effect of feature type and other interactions[$F(1, 22) < 1$, $p > .34$]. Simple main effect analysis showed that the difference between moving durations was significant for identical condition [$p < .001$], but was not significant for unique condition [$p = .160$].

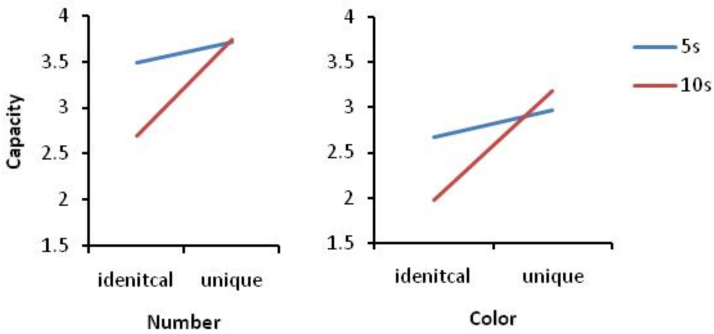

Fig. 4. Identification capacity as a function of feature type, object differentiability and moving duration

4 General Discussion

In the study, we investigated the human observers' abilities of tracking multiple moving items and maintaining their identities. We found that observers can, to some extent, maintain the identities of multiple objects. However, observers' capacity of ID task was dependent on feature type. ID capacity is about 3.6 items for digit-based representation, and about 3 items for color-based representation. Thus we suggest that the human capacity of maintaining multiple moving object identities is about three to four items.

4.1 The Unique Benefit

The results showed that tracking capacity decreased with moving duration, but only with identical items. The use of unique items improved the general tracking performance and eliminated the difficulty caused by long moving duration. This was contrast to the lack of unique advantage in [4]. The main difference between their research and this study is an additional identification task may stimulate participants to adopt a voluntary strategy to encode the distinct feature. Participants might have attempted to encode the unique feature to help them with the tracking task. This may suggest that tracking implemented by early vision is "feature-blind" as contended by FINST theory [5], but the object identity differentiated by visual properties can be encoded or accessible from higher level cognitive processes when a voluntary strategy involved.

It also showed that ID capacity decreased with moving duration, but only with identical items. The use of unique items improved the general ID performance and eliminated the difficulty caused by long moving duration. This was consistent with intuitive inference and had collectively demonstrated content addressable representations in multiple-object tracking [3]. The most important result was that ID capacity was better with number labels than with color labels no matter the labels were shown at the initial static phase or during the movement, but this was not the case for tracking capacity. It indicated a greater role of digit than color labels for content addressing. It may due to the more significant distinguishability of digit than color. We can distinguish and remember 8 digits more easily and quickly than 8 colors and thus require less resource to process these features. In fact, the digit span of adult is greater than the color span (6 items, [6]). A less resource-demanding process of identity-related feature would lead to more effective improvement on tracking.

4.2 Implications and Applications of the Present Study

In this study, we explored the capacity of multiple object tracking and identification of human observers. We found that human capacity of keeping track of object identities is about three to four items. This limited capacity should be taken into account in the computer vision system involving human monitoring. In a human-computer interaction, for example, to ensure efficient monitoring, at most four targets should be displayed to a human surveillant. And since identity-related features improve human identification performance, it is helpful to label every object in the display with unique features with less resource-demanding.

We also found that identification of multiple moving objects is challenging for human observers just as it is for computer vision systems. However, for human observers, this performance can be improved by keeping the identity-related featural information of all items visible during their movement, which suggested that human can use feature information to maintain the identity of multiple moving objects. This featural correspondence facilitation could be applied in vision system by creating a feature space to map the identity of multiple objects. However, different feature don't have the same impact on identification. This provides potential implications for computer MOT. Simple features are not selected arbitrarily to serve as efficient cues for improving identification during tracking.

Furthermore, our data showed different impacts of unique benefit on tracking and ID performances, and indicated two separate mechanisms of resource allocation. For computer vision applications, tracking and ID are both probably involved in MOT. The separate mechanisms may imply a general principle of design with two vision systems (or modules) of MOT in computer vision application.

Acknowledgements

This research was supported in part by grants from National Basic Research Program of China (2006CB303101), the National Natural Science Foundation of China (30600182, and 30500157) and Project for Young Scientists Fund, Institute of Psychology, Chinese Academy of Sciences (08CX032003).

References

1. Pylyshyn, Z.W., Storm, R.W.: Tracking multiple independent targets, evidence for a parallel tracking mechanism. Spatial Vision 3, 179–197 (1988)
2. Bahrami, B.: Object property encoding and change blindness in multiple object tracking. Visual cognition 10(8), 949–963 (2003)

3. Horowitz, T.S., Klieger, S.B., Fencsik, D.E., Yang, K.K., Alvarez, G.A., Wolfe, J.Y.: Tracking unique objects. Perception & Psychophysics 69, 172–184 (2007)
4. Klieger, S.B., Horowitz, T.S., Wolfe, J.M.: Is multiple object tracking colorblind? Journal of Vision 4(8), 363 (2004)
5. Pylyshyn, Z.W.: Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. Visual cognition 11(7), 801–822 (2004)
6. Zoelch, C., Seitz, K., Schumann-Hengsteler, R.: From rag(bag)s to riches: measuring the developing central executive. In: Schneider, W., Schumann-Hengsteler, R., Sodian, B. (eds.) Young Children's Cognitive Development: Interrelationships among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind. Erlbaum, Mahwah (2005)