Decomposing the spatiotemporal signature in dynamic 3D object recognition

Ying Wang

Institute of Psychology, Chinese Academy of Sciences,
Beijing, China, &
Graduate University, Chinese Academy of Sciences,
Beijing, China

Kan Zhang

Institute of Psychology, Chinese Academy of Sciences,
Beijing, China

The current study investigated the long-term representation of spatiotemporal signature (J. V. Stone, 1998) and its coding nature in a dynamic object recognition task. In Experiment 1, the observers' recognition performance was impaired by an overall reversal of the studied objects' learning view sequences even when they were unsmooth, suggesting that the spatiotemporal appearance of the objects was used for recognition, and this effect was not restricted to smooth motion condition. In another four experiments, a feature reversal paradigm was applied that only the global-scale or local-scale dynamic feature of the view sequences was reversed at a time. The reversal effect still held, but it was selective to the sequence's feature saliency, suggesting that statistical representation based on specific features instead of the whole view sequence was used for recognition. Furthermore, top-down regulation on sequence smoothness was observed that the observers perceived the objects as moving in a smoother manner than they actually were. These results extend an emerging framework that argues the spatiotemporal appearance of a dynamic object contributes to its recognition. The spatiotemporal signature might be coded in a feature-based manner under the law of perceptual organization, and the coding process is adaptive to variation of the sequence's temporal order.

Keywords: dynamic object recognition, reversal effect, spatiotemporal signature, perceptual organization, top-down regulation

Citation: Wang, Y., & Zhang, K. (2010). Decomposing the spatiotemporal signature in dynamic 3D object recognition. *Journal of Vision, 10*(10):23, 1–16, http://www.journalofvision.org/content/10/10/23, doi:10.1167/10.10.23.

Introduction

Recognizing dynamic objects is a computational challenge though natural and effortless to human beings. As the relative positions of observers and objects change continuously, so do the projected retinal images. How does the brain recognize a dynamic object from the changing view sequence? A theoretical solution is to perceive the object as multiple static view images and then extract from each single view the shape information of the object, like the descriptions of its 3D structure (Biederman, 1987; Hummel & Biederman, 1992; Marr & Nishihara, 1978), multiple represented views (Bülthoff & Edelman, 1992; Poggio & Edelman, 1990; Riesenhuber & Poggio, 2000), or their combination (Foster & Gilson, 2002; Tarr, Bülthoff, Zabinski, & Blanz, 1997) to accomplish successful recognition. This image-based method is not only a theoretical hypothesis from the view of static object recognition research. It has gained empirical support from several biological motion studies (Beintema & Lappe, 2002; Bertenthal & Pinto, 1994; Lange, Georg, & Lappe, 2006; Reid, Brooks, Blair, & van der Zwan, 2009). For example, Reid et al. (2009) have shown that a snapshot with only a few dots (from a dynamic view sequence) alone was sufficient to give rise to the perception of a point-light walkers' walking direction, implicating that the static images might convey both shape and motion information of a point-light walker so as to fulfill the recognition.

Although the above theories provide a self-sufficient method for the recognition of dynamic objects, it might still be an incomplete framework for the lack of interpretation on the role of motion. As suggested by recent neurophysiologic studies, motion might be integrated into the framework of dynamic object recognition with shape through the interaction of dorsal and ventral visual pathways (Farivar, Blanke, & Chaudhuri, 2009; O'Toole, Roark, & Abdi, 2002; Sarkheil, Vuong, Bülthoff, & Noppeney, 2008; Schultz, Chuang, & Vuong, 2008). Accumulating behavioral evidence also reveals that the motion pattern of an object might support its recognition through different mechanisms. It has long been observed that people can perceive the 3D structure of an object by watching its 2D projection in motion (Farivar et al., 2009; Siegel & Andersen, 1988; Ullman, 1979; Wallach & O'Connell, 1953). Learning a small range of views through apparent motion enhances the effect of view interpolation, i.e., beneficiation of non-studied views that are between the studied views over the non-studied views

doi: 10.1167/10.10.23 Received December 8, 2009; published August 24, 2010 ISSN 1534-7362 © ARVO

that precede or follow the studied trajectory (Friedman, Vuong, & Spetch, 2010, 2009; Kourtzi & Shiffrar, 1997; Spetch & Friedman, 2003), as well as facilitates view generalization to the post-trajectory views compared with the preceding extrapolated views (Friedman et al., 2010, 2009; Vuong & Tarr, 2004). Besides a facilitator for shape processing, motion is also considered to be an independent cue for the recognition of walking people's identities (e.g., Cutting & Kozlowski, 1977; Richardson & Johnston, 2005) and other non-biological dynamic objects (e.g., Newell, Wallraven, & Huber, 2004; Setti & Newell, 2009).

Although the growing evidence suggests the distinctive roles of shape and motion in human dynamic object recognition, relatively little is known about the spatiotemporal coding of the dynamic objects, especially how object's motion is represented for recognition. Theoretically, there are at least three alternative ways for the coding of dynamic objects in the recognition task: (1) only multiple view images but no object motion is coded; (2) the whole view image sequence with specific sequence order is coded; (3) the view image sequence is coded as a statistical representation of some spatiotemporal features instead of a replication of the physical sequence.

The first hypothesis would predict that a change to an object's motion but not views should have no effect on recognition. However, results from motion reversal studies indicate otherwise. In Stone's (1998, 1999) studies, reversing the direction of depth rotation impaired the recognition of two novel 3D objects when the motion directions were characteristic for the objects during learning. The reversal effect has been observed for both rigid (Liu & Cooper, 2003; Stone, 1998, 1999; Vuong & Tarr, 2006) and non-rigid (Chuang, Vuong, Thornton, & Bülthoff, 2006) motions, using both explicit and implicit tasks (Liu & Cooper, 2003), and among objects of more or less distinctive features (Spetch, Friedman, & Vuong, 2006; Vuong & Tarr, 2006) or of biological significance (Hill & Johnston, 2001). Since the reversing manipulation only changes the temporal order of the view images without any shape information loss, it provides strong evidence that the spatiotemporal appearance of the objects, namely spatiotemporal signature (Stone, 1998), has been used for recognition. Despite ruling out the image only hypothesis, we cannot disentangle the latter two hypotheses according to our knowledge; though there are several mathematical methods to transform the view sequence into features that are applicable for machine recognition (Casile & Giese, 2005; Troje, 2002), raising the possibility that humans might use similar strategies in dynamic object recognition. Surprisingly, we have poor evidence about whether and how humans decompose the dynamic object into meaningful spatiotemporal elements.

Another important issue for dynamic object coding is the perceptual constraints on the coding process. One of these potential constraints is the smoothness of the spatiotemporal sequence, i.e., the spatiotemporal continuity of the view images' sequence. Since smoothness is a basic constraint for the computation of optic flow (Horn & Schunck, 1981), it is natural to ask whether it would have direct impact on the coding of spatiotemporal signature. There is evidence that smoothness modulates the facilitating effect of motion on shape processing. View association was only learnt when the motion sequence was spatiotemporally smooth rather than random (Balas & Sinha, 2008; Wallis & Bülthoff, 2001). The same restriction was found for the view generalization advantage of post-trajectory views compared with pre-trajectory views, while the generalization to interpolated views was not restricted to but still gained benefit from smooth sequence (Friedman et al., 2010, 2009; Lawson, Humphreys, & Watson, 1994). Nevertheless, breaking down the sequence's smoothness was not necessary to impair the representation of the static views directly (Harman & Humphrey, 1999; Liu, 2007). Despite the controversial results, the role of sequence's smoothness on the coding of spatiotemporal signature has not been explicitly tested in the previous dynamic object recognition studies.

In the present study, we were interested in the way the dynamic object is being represented, especially to answer the question whether the representation learnt from the spatiotemporal signature was decomposable to some extent. We also explored the potential interaction between the representation and its coding constraints, particularly whether the coding of spatiotemporal signature was restricted to smooth sequence.

Our basic assumption was that an ecologically feasible representation of dynamic objects should be subject to both perceptual organization (decomposable sequence) and top-down modulation (regulated sequence). Accordingly, we would predict the view sequence to be coded in an organized way rather than as an exact replication of the original sequence. We would also predict that the observers have tolerance to the unsmooth sequence as long as the sequence has statistically salient features. There are three reasons behind these hypotheses. First, we could not exclude the analogy that humans extract some features from a dynamic object as suggested from the view of computational vision (Casile & Giese, 2005; Troje, 2002). Second, according to the Gestalt psychology (Koffka, 1999) and some recent studies (Blake & Lee, 2005; Gepshtein & Kubovy, 2000), the perception of spatiotemporal stimuli is regulated by several laws of visual organization in both spatial and temporal dimensions (e.g., common fate). In other words, spatiotemporal stimuli might be organized according to their own physical appearances. Third, the remarkable ability of spatial regulation in dynamic object recognition (Bülthoff, Bülthoff, & Sinha, 1998; Sinha & Poggio, 1996) led to the conjecture that similar top-down regulation might modulate the perception of temporally unsmooth sequence for the purpose of recognition. In the current study, we tried to look for evidence in support of the assumptions using a modified reversal paradigm and object recognition task.

The typical reversal paradigm (Stone, 1998) had a learning phase and a test phase. The observers' task was to discriminate two dynamic objects. In the learning phase, two objects always moved in different characteristic trajectories (usually reverse to each other), so they were distinguishable in both shape and motion. In the following test phase, the two objects appeared in either the original motion trajectory, or in a reverse trajectory. If the object's motion was used for recognition, then the reverse condition might have worse recognition performance than the studied motion condition.

In Experiment 1, we examined whether the coding of spatiotemporal signature was restricted by sequence's smoothness with partially scrambled view sequence (see Methods section in Experiment 1 for the details). Previous studies regarding spatiotemporal smoothness (Balas & Sinha, 2008; Friedman et al., 2010, 2009; Harman & Humphrey, 1999; Lawson et al., 1994; Liu, 2007; Wallis & Bülthoff, 2001) did not directly test its effect on dynamic object recognition for they usually included only static views during the test, which might hinder the expression of dynamic information even if it had been represented. The reversal paradigm used by the current study allowed us to examine the role of sequence's smoothness in both learning and recognition. If smoothness was a constraint for the coding of spatiotemporal signature, we would expect to observe no reversal effect when the learning and test sequences were both unsmooth. Whereas, if the observers had a tendency to regulate the partially scrambled sequences for the coding of spatiotemporal signature as we assumed, we would expect to observe the reversal effect for the unsmooth sequences.

In another four experiments, we tried to scrutinize the representation of spatiotemporal signature with a feature-reversal paradigm. We reversed dynamic features of different temporal scales for each sequence, either at a local (Experiments 2a and 3a) or global (Experiments 2b and 3b) level, to see if the observers would use these features for the purpose of recognition. The rationale was if the global and local dynamic features of the view sequence was changed, respectively, and the observers were only sensitive to the reversal of one dynamic feature but was not to another for the same sequence, then the observers might not represent the whole sequence, instead, they might use the representation of a specific feature for recognition. We also manipulated the dynamic feature's saliency of the object's view sequence (Experiments 2a and 2b versus Experiments 3a and 3b) to see its effect on the coding process. If the feature used for recognition was related to its saliency in a specific view sequence, then the coding process might be impacted by the availability of the dynamic feature.

In all the experiments, we introduced the partially scrambled sequences, which only had the global feature scrambled (Experiments 1, 2a and 2b) or the local feature scrambled (Experiments 3a and 3b) based on the smooth view sequence of the objects. There were four concerns for the use of these partially scrambled sequences: (1) they served as our tool to test whether spatiotemporal smoothness was a constraint for the use of characteristic motion in recognition. (2) For the partially scrambled sequences, we could reverse their global or local feature, respectively, without changing their physical difficulties, though it was impossible for the smooth sequence or the totally random sequence. (3) All the objects we used had salient form and we ensured that they were easy to discriminate even when presented in our scrambled sequences (which was verified by the data), although the absolute quality of shape representation was not within the main focus of the current study. (4) The sequences were not totally scrambled, which left them physically describable, so we had the chance to see whether the non-scrambled (salient) features would be used for recognition as we assumed.

Experiment 1

In Experiment 1, we examined the coding of spatiotemporal signature under the constraint of sequence's smoothness. We used objects rotating in partially scrambled sequences to see if their motion patterns would be learnt for recognition. We also manipulated the repetitiveness of the learning sequence to test the exposure intensity needed for a scrambled view sequence to be remembered. In a previous work by Liu and Cooper (2003), they found that the observers' old–new recognition performance was impaired by motion reversal even though they had viewed each object rotating only once. This led to a conjecture that the use of motion cue is automatic. We tested this conjecture under the unsmooth motion condition by keeping the objects' view sequence order either fixed or unfixed to see its effect on recognition.

Methods

Participants

Twenty-six college students (14 females, average age 22) volunteered to take part in the experiment for monetary payment.

Stimuli

The objects we used were created by 3D Studio Max 6.0 and rendered in gray texture (Figure 1). They were assembled by equal-sized cubes based on the rules used by Gauthier et al. (2002) and Tarr (1995). Six objects were assigned to two groups for the counterbalancing of experimental conditions. Considered that the sense of similarity among these shapes may be based on quite

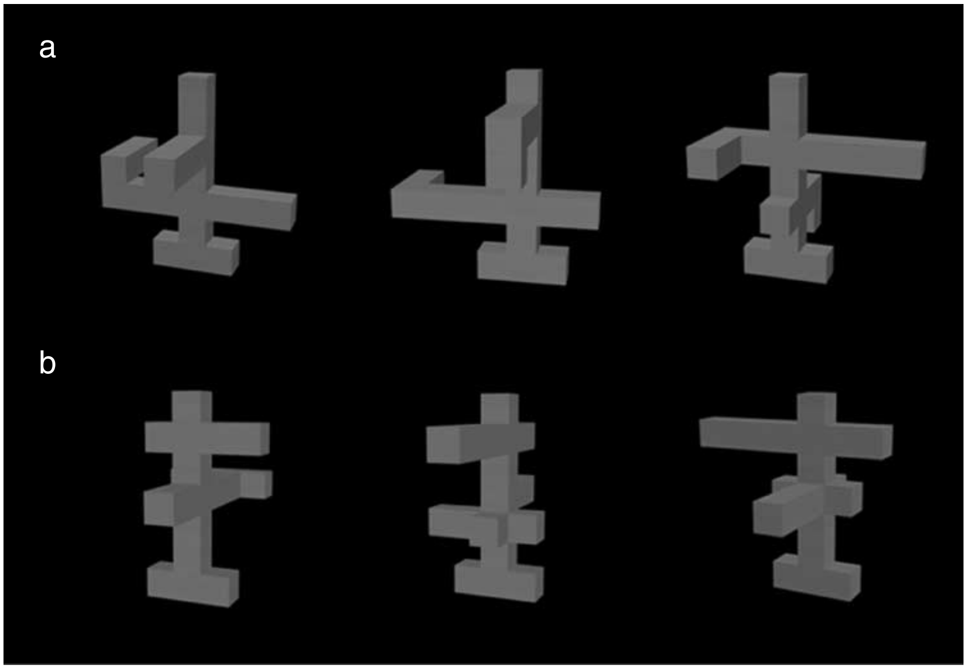


Figure 1. Three-dimensional objects (upper row: group a; lower row: group b) used in all five experiments.

different criteria for each individual (which was suggested by a pilot study), we did not match the difficulties of the objects intentionally. However, the pilot study showed that there was no significant difference in recognition performance between these two object groups.

For each experimental block, a pair of reversed scrambled trajectories was created by three steps.

First, 30 views were generated to simulate a smooth apparent motion trajectory of an object rotating around 360 degrees with equal view intervals of 12 degrees. The speed of motion was 120 degrees per second, with a full motion cycle of 3 s. These views were arranged in either clockwise or counterclockwise order relative to a predefined axis (Figures 2a and 2b).

Then the whole sequence was divided into 10 ordered chunks, each depicting a rotation trajectory of 36 degrees. To get scrambled, the displayed order of the ten chunks was randomized, while the view sequence within each chunk remained unchanged (Figures 2b and 2c).

After that, a 1–30 to 30–1 reversal was conducted to the obtained view sequence to generate a completely reverse trajectory. The two trajectories were then randomly assigned to the two studied objects in a learning block (see Movie 1).

The sequences generated in this way were always jittering as they typically jumped among non-adjacent chunks. However, since we only randomized the sequence partially (on chunk level), the apparent motion was preserved within each chunk.

All the experiments were programmed using the platform of ImageTcl (Owen, Tang, & Xiao, 2003). The stimuli were displayed on the center of a 17-inch LCD in a dark room. The maximum height of the object on the screen is about 13 cm. A semitransparent gray sphere, a bit larger than the stimuli, was always shown before a trial began as an attention guide.

Design

We manipulated (1) repetitiveness of the sequence (fixed, unfixed), (2) motion change of the test sequence (original, reverse, novel), and (3) object's type (studied, novel) as within-participant factors. Because it was the first time we introduced the scrambled sequence into the reversal paradigm, we used the novel trajectory to test the participants' sensitivity to the studied trajectory, as well as to verify the validity of our manipulation on sequence dynamics. Additionally, we added a non-studied third object to the test trials without informing the participants to test their sensitivity to the change of objects' shape. Since our task was object discrimination, we wanted to check if our participants really had some memories of the studied shapes.

Each participant completed two blocks, with either fixed view sequence or unfixed view sequences. The order of these two blocks and the corresponding object groups were all counterbalanced among participants. Two objects were randomly selected from the object group as studied objects for each participant within each experimental block. In the fixed sequence condition, the sequence order of a particular object kept the same over all the learning trials for a participant. While in the unfixed sequence condition, the sequence order of each object altered from trial to trial. For both conditions, the sequence matrix of learning trials was totally reverse to each other for the two objects. During the test phase, there were three kinds of motion sequences for the studied object: either the same

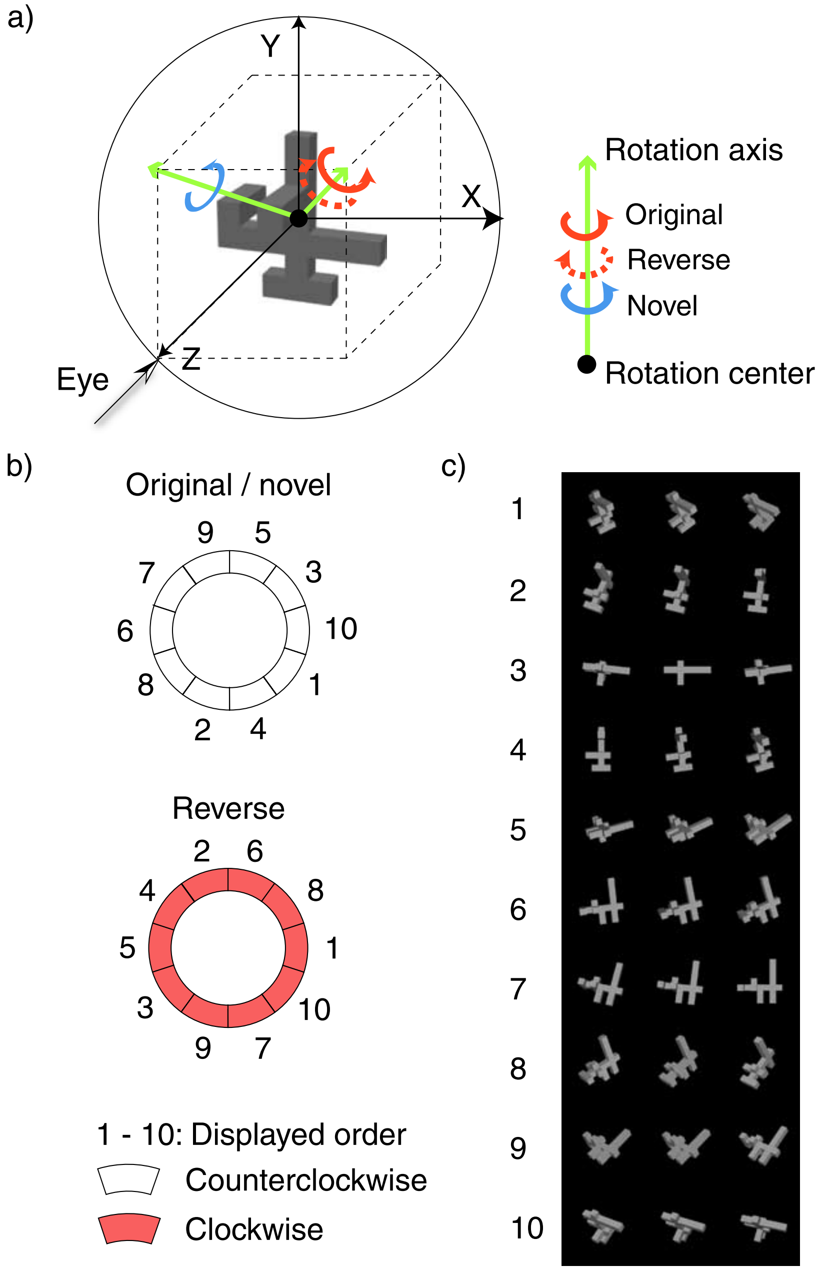
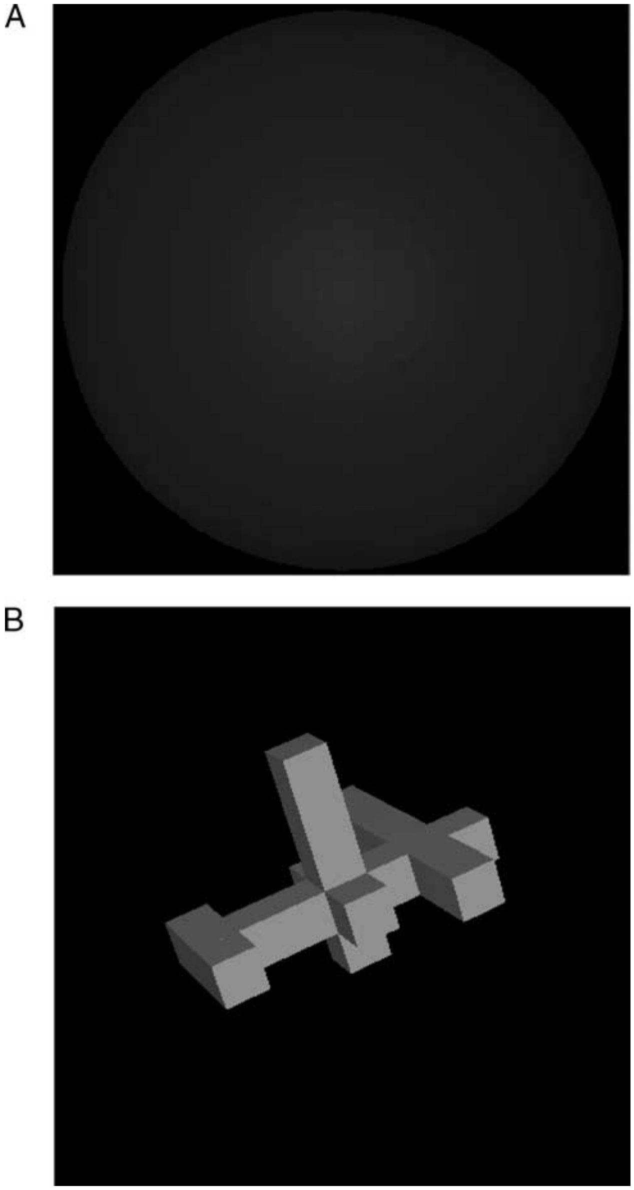


Figure 2. Illustration of the partially scrambled sequences used in Experiment 1. (a) Axes and directions of the motion trajectories disregarding the scrambling manipulation. The original and the reverse trajectories both rotate around the axis of $x = y = z$ (which is defined in the viewer-centered coordinate, X to the right, Y up, Z to the eyes, and getting through the center of the object) but in different directions. The novel trajectory rotates around the axis of $-x = y = z$ and inherits all of the other parameters of the original trajectory. (b) Sequence dynamics of different trajectories disregarding the rotation axes. Each ring indicates the 30 continuous views (in 10 ordered chunks, separated by the black lines) around the object for 360 degrees. Numbers around each ring indicate the DISPLAY ORDER of the ten chunks, from 1 to 10 as a cycle, starting from each of the ten chunks in equal probability. The upper ring illustrates a scrambled sequence that was displayed in the chunk order from 1 to 10, indicating by the number. The sequence is jittering between adjacent chunks. For example, the 1st displayed chunk is 36 degrees away from the 2nd. The lower ring illustrates a sequence that is reverse to the upper one. The correspondence between the two sequences is: the 1st chunk in the reverse sequence is the 10th in the original sequence, 2nd to 9th, 3rd to 8th, and so on. The red and white colors indicate the view sequence order within a chunk. In Experiment 1, all local chunks within a single trajectory have the same rotation direction (same color), counterclockwise or clockwise. The two rings have reverse chunk order (global dynamics) as well as reverse view sequence within each chunk (local dynamics), so the two sequences were completely reverse to each other. (c) A specific object moving in the view sequence illustrated by the upper ring in (b). Each row depicts a chunk of three views, displayed in the order from left to right. The correspondence between the ten chunks in (b) and (c) was indicated by the numbers (1–1, 2–2, …, 10–10).



Movie 1. Demonstration of the stimuli used in Experiment 1. The view sequences were globally scrambled (randomized chunk order). Two objects moved in totally reverse trajectories.

as its learning sequence, a completely reverse sequence, or a totally novel sequence, which inherited all the motion parameters of its learnt trajectories except a new rotation axis (Figure 2). The novel object also had three kinds of test trajectories, two of which were the same as the two studied objects' trajectories and the other one was the same as the novel trajectory of one studied object.

Procedure

Before each experimental block, the participants were told that they were going to learn two dynamic objects displayed in animation, and their task was to remember the two objects and press the corresponding key to each one. Each participant would complete two blocks with either fixed or unfixed sequences. However, they were not told about the difference, neither about the scrambling manipulation of the sequences.

Each block began with a familiarizing phase; the participants were asked to study two animated objects and their predefined codes. Each object appeared once moving in two intact motion cycles. After it disappeared, a number would appear on the screen to indicate its code. The participants should press the corresponding number key "1" or "2" to make the number disappear. At this stage, they got an impression of the animation and its code, while getting familiar with the key-pressing task.

In the learning phase, participants learnt to discriminate the two objects with feedback. They were instructed to press the number key as accurately and as quickly as possible thereafter a piece of animation disappeared, with specific emphasis that they had to observe each of the two objects carefully and remember them for a later test. In a typical trial, one object moved in a full cycle and then disappeared. The participant pressed a key and got a feedback on the screen. The learning phase had two successive sections. Each section contained twenty trials, ten for each object, in totally randomized order. For the fixed condition, each object always appeared in the same trajectory, with equal probability to start from each chunk (so different initial view for each of the ten trials). For the unfixed condition, the sequence order was regenerated for each trial thus with randomized initial views either. The second section always replicated the trials in the first one but in a different randomized trial order. This manipulation avoided biases toward certain views caused by either the primacy effect or the recency effect of the animations.

After feedback learning, the participants were instructed to do the discriminating task in a test phase without feedback. They were told that they might see something different in the animations (without the details) and they had to make judgments based on their memories. Another difference from the learning phase was that the participants did not need to wait for the object to disappear to respond. Three objects (two studied, one novel) moved either in the two studied trajectories or in a novel trajectory, repeated ten times for each combination starting from different initial views as in the learning phase. Each participant made 90 discrimination judgments in a test block with randomized trial order. During the experiment, the participants initialized trials themselves by pressing the "space" key. So they had self-timing breaks in between trials.

Results

To ensure the participants had learnt the two objects proficiently, we set up a 70% mean accuracy criterion to the feedback learning phase. Participants had to meet the criteria to be included in the analysis. To reduce the interference of extreme data caused by other accidental errors, we refined the data in the test phase based on two rules. First, trials with reaction times beyond the range of 3 standard deviations higher or lower than individual means were excluded. Second, we conducted a by-object screening to avoid the case that one had mixed up memory of the studied objects in the test. If a participant had no more than 30% correct response to a studied object moved in original motion, his or her responses to this object were excluded from analysis. All the experiments in this study refined data following the above rules. Two participants failed to enroll in the analysis and another 165 trials (3.82% of the total) were excluded in Experiment 1. We report results based on the valid data.

Mean accuracies in the learning phase for the fixed and unfixed conditions were 93.7% and 90.6%, respectively. The mean reaction times of correct trials were 555.5 ms and 570.53 ms. Paired sample t-test revealed no difference between the two learning conditions in both accuracy ($t(23) = -0.443$, $p = 0.662$) and RT data ($t(23) = 1.002$, $p = 0.327$).

To access our interests on the repetitiveness of view sequence (fixed, unfixed) and the change of test motion sequence (original, reverse, novel), we carried out a two-way repeated measures ANOVA on both accuracy and RT data of the studied objects in test. A main effect of test motion sequence was observed on both accuracy ($F(2, 46) = 7.627$, $p = 0.001$) and RT ($F(2, 38) = 6.817$, $p = 0.003$). No significant effect of sequence repetitiveness was found. However, the trend was steady among all the test motion conditions (Figure 3). Recognizing objects moved in the fixed condition was a bit easier and faster than in the unfixed condition, though it did not reach statistical significance. There was no interaction between the repetitiveness of learning sequence and the change of test sequence.

We conducted planned comparison tests to examine the effect of motion trajectory under each learning condition. The results showed that participants made more errors when judging the objects in a reverse trajectory. This trend was reliable for both fixed learning condition ($t(23) = -2.114$, $p = 0.046$) and unfixed learning condition ($t(23) = -2.742$, $p = 0.012$). Observing objects moved in a novel trajectory also impaired performances for both fixed ($t(23) = -2.524$, $p = 0.019$) and unfixed ($t(23) = -2.905$, $p = 0.008$) conditions. For RT data, the patterns were similar, though the significant delay in response was only observed between novel and original trajectories ($t(22) = 2.064$, $p = 0.051$ for fixed condition; $t(23) = -3.732$, $p = 0.001$ for unfixed condition).

For the novel objects, the participants were not explicitly asked to press a correspondent key except the wrong keys (1 and 2), so there is no correct response. However, if the participants were sensitive to the shape change, their reaction time to these novel objects should be longer than that to the studied objects. This assumption was verified by the three-way repeated ANOVA (learning condition, test sequence, and object's type). The main effect of object's type was significant ($F(1, 19) = 16.109$, $p = 0.001$) and the robust response delay for the novel objects was found for all the six conditions combined by the learning and test sequences (either at 0.05 or 0.01 level).

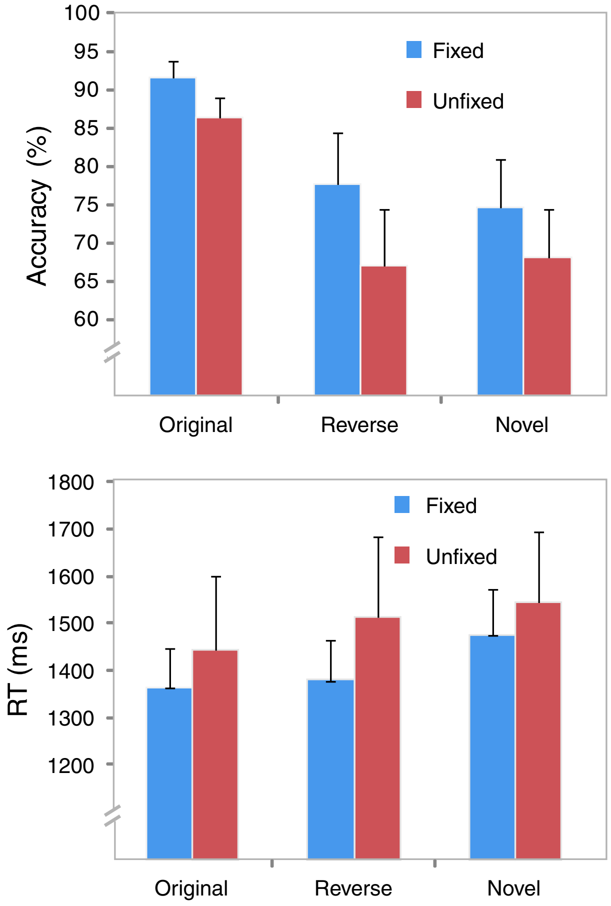


Figure 3. Accuracy and RT of responses to studied objects as a function of view sequence repetitiveness (fixed, unfixed) and test motion trajectory (original, reverse, novel). Error bar indicates one *SE*.

Discussion

In Experiment 1, we found an interesting motion reversal effect that was not restricted to spatiotemporally smooth sequence. The decrease of performance in the reverse and novel motion conditions revealed the observers' perceptual sensitivity to the physically scrambled sequences, as well as the role of spatiotemporal signature in recognition.

Since it was the first time we got such a reversal effect using scrambled sequence, we would like to rule out the possibility that this result was simply a perceptual artifact. The best discrimination rate in the fixed-original condition was up to 92%, which was comparable to the learning baseline. Meanwhile, even the worse performance in the unfixed novel condition was much higher than the chance level (0.68 vs. 0.5). These data indicated that the manipulation we applied on sequence dynamics was perceptually adaptable. The observers were able to learn the novel 3D objects in physically scrambled sequences and generalize to the novel views in unstudied motion patterns. Additionally, lack of interaction between the fixed and unfixed conditions also suggested that the reversal effect we got was a general phenomenon that was not restricted to a specific perceptual manipulation. The observers were not only able to discriminate object identities from the scrambled sequences, their judgments were biased by the test sequence. In this sense, we claimed that the observers used the spatiotemporal signature for recognition and the coding process was not very sensitive to the sequence's smoothness, at least under the current manipulation. We basically replicated the findings of previous studies (Liu & Cooper, 2003; Stone, 1998, 1999; Vuong & Tarr, 2006) and extended the working condition of characteristic motion as a cue for recognition.

The main effect of repetitive exposure was not found, though we noticed that the advantage of the fixed sequence over the unfixed sequence was quite steady among all the test conditions in both accuracy and RT. This trend was also consistent with the performance in the learning phase. It might bear additional evidence to the notion that dynamic object recognition was not totally view-based. Since the participants were exposed to the same set of views under the two learning conditions, their different performance in these two conditions (if not accidental, as we assumed) could only be attributed to the difference in motion sequence and was reasonable if explained in the framework of temporal associative learning (Miyashita & Chang, 1988). If repeated learning was needed to bind different views into view-invariant representation, then we should predict a representation's quality being correlated with the binding intensity. As in our study, the binding intensity of views in the fixed learning condition was larger than that in the unfixed learning condition, as well as the quality of representation that was indicated by the performance. Though some previous studies had restricted view binding as a specific function of smooth sequence (Balas & Sinha, 2008; Wallis & Bülthoff, 2001), their conclusions were not totally comparable to ours because they did not manipulate the binding intensity of the learning sequence directly. There was still a possibility that the partially scrambled sequence contributed to view binding in a weaker form than the smooth sequence, but this effect might still exist and might get benefit from repetitive exposure. Though not being verified in the present study, it would be an interesting hypothesis to test in the future.

Another intriguing finding of Experiment 1 was that even under the unfixed learning condition, reversing motion trajectory did impair recognition performance. This elicited our thinking on the question of how the scrambled sequences were represented. An intuitive explanation was that the observers might really have "copied" the whole scrambled sequences to their minds, no matter the scrambled sequences were fixed or unfixed during learning. As in Liu's (2007) study, the observers showed an implicit memory of the object's trajectory in an old–new recognition task even when they had only learnt it once. However, whether this capacity could be extended to the scrambled sequence and discrimination task in the current study was still unclear. As we noted above, the temporal binding effect under unsmooth context should be weak even there was any. It was doubtful that the observers had remembered the orders of all the scrambled sequences under the low exposure intensity. An alternative explanation was that the observers remembered some features of the sequences, and their memories of these features were consolidated even under the unfixed condition. In the current experiment, though the chunk order (global dynamic feature) was always changing in the unfixed learning condition, some other feature of the sequence, such as the sequence order within each chunk (local dynamic feature), was never changed. Notice that in this experiment all the chunks had a "common fate" (counterclockwise or clockwise, see Figure 2) among all the learning trials for a specific object. If the participants used this consistent local dynamic feature to represent the sequence, they might have gained knowledge about the object's spatiotemporal appearance through repetition even in the unfixed condition. Experiments 2a and 2b were designed to examine these two explanations and further explore the possible representation of the spatiotemporal signature.

Experiments 2a and 2b

The goal of this experiment was to test whether people had remembered the whole globally scrambled sequence or they just used representation of some dynamic features, such as the temporal order within the chunks, for recognition. We modified the temporal manipulation of the reversal paradigm to get a subtle look into the representation of global and local dynamic features, which was defined by the sequence's temporal order in the current study, of a sequence, respectively.

Methods

Participants

Twenty-seven college students (13 females, average age 23) were recruited for Experiment 2a and 20 students (half females, average age 23) for Experiment 2b. They were all paid for participation and did not take part in the previous experiment.

Design and procedure

The overall design and procedure were similar to Experiment 1, with several exceptions. In the learning and test phases within each block, the sequence order for a certain object was always fixed. In the test phase, each studied object moved in either the original trajectory or a reverse trajectory (locally reverse in Experiment 2a, and globally reverse in Experiment 2b), initiating randomly from each of the ten chunks, though we did not test the novel trajectory in these two experiments. We preserved the non-informed novel object to test the participants' sensitivity to shape. Each participant completed two blocks, which differed only in the objects employed.

The classical reversal paradigm was modified to dissociate the effects of global and local dynamic features. In Experiment 2a, we reversed the local dynamic feature (sequence order within each chunk) while keeping the global dynamic feature (chunk order) of the sequence unchanged. In Experiment 2b, the local feature was kept the same, while the global feature was reversed (Figure 4). These kinds of manipulations made it possible to get sequences reversed at either global or local level without causing change in physical difficulty or the feeling of smoothness. The motion trajectories of the object pair in a learning block differed only in global or local dynamics, according to the experimental conditions. If the participants used the whole sequence for recognition, we would expect to observe reversal effects in both Experiments 2a and 2b. Otherwise, if the participants were only sensitive to the constant local dynamic feature, the impairment on recognition performance should only occur when the local feature was reversed (Experiment 2a).

Results

In Experiment 2a, three participants failed to pass the criteria. Another 60 trials (2.08% of total) were removed based on the rules we described in Experiment 1. In Experiment 2b, 100 trials (4.17% of total) were rejected for analysis.

Mean accuracies of feedback learning for Experiments 2a and 2b were both 90.8%. Reaction times were also quite near. The data were 633 ms and 613.4 ms, respectively.

The intriguing results came from the test phase. We found that reversing the temporal order of each chunk (locally reverse) did impair the discrimination accuracy severely ($t(23) = 1.881$, $p = 0.036$, one-tailed) from 91.7% ($SE = 0.02$) to 78.9% ($SE = 0.066$) in Experiment 2a. Response to the original motion condition (1573.8 ms, $SE = 94.06$) was slightly faster ($t(23) = 1.271$, $p = 0.109$, one-tailed) than to the locally reverse condition (1630.6 ms, $SE = 108.33$). Compared with the effect of local reversal, reversing the scrambled chunk order (globally reverse) almost had no influence on the discrimination task in Experiment 2b. There was no difference among accuracies ($t(19) = 0.433$, $p = 0.335$, one-tailed) of the original motion condition (90%, $SE = 0.02$) and the globally reverse condition (89.4%, $SE = 0.021$). The RT data neither showed any reversal effect. Actually there was a bit trade off. Reaction in the reverse condition (1376.1 ms, $SE = 76.91$) even turned out to be a bit faster ($t(19) =$

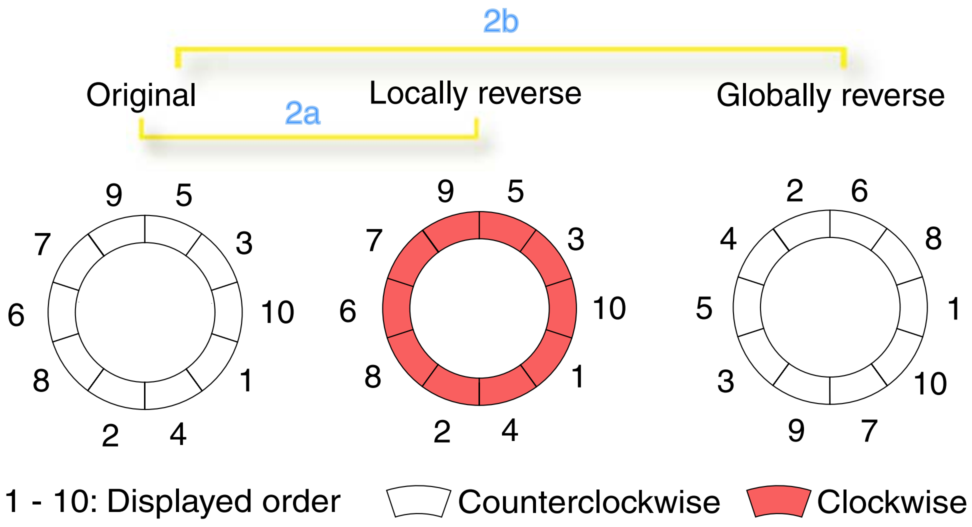


Figure 4. Design of Experiments 2a and 2b. The sequences were scrambled in the same way as that in Experiment 1. They were scrambled at the global level (randomized chunk order), indicated by the numbers, but with salient local feature (consistent rotation direction within each chunk), indicated by the colors. In Experiment 2a, view sequence order of each local chunk (colors) was reversed; in Experiment 2b, the global chunk order (numbers) was reversed.

1.991, p = 0.062) than that in the original motion condition (1440.5 ms, SE = 92.75).

Discussion

Back to the question raised by Experiment 1, we ruled out the possibility that the observers used the whole view sequence for recognition. Actually, they were quite insensitive to the scrambled chunk order (global reversal, Experiment 2b), while dramatically affected by the reversal of the temporal order within each chunk (local reversal, Experiment 2a).

This intriguing new finding implied that the spatiotemporal signature of a dynamic object might be decomposed into local feature that represented the local dynamics of the sequence, which then served as an available cue for recognition. This might explain why repetitive exposure did not have any influence on the reversal effect in Experiment 1. The complete reversal paradigm reversed both the global and local features of a sequence. So if people only had extracted local feature from the scrambled sequences, it would not matter whether the global order was fixed or unfixed during learning.

It seemed that the physically scrambled sequences appeared to be perceptually continuous as long as the local chunks had consistent order. Most of the participants reported that they saw the object rotating in an orderly manner in the post-experimental interview. This result supported a top-down “recognition-before-perception” effect on dynamic object perception, which was modulated by learning (Bülthoff et al., 1998).

The seeming unimportance of global order shown by Experiment 2b might have various explanations. It could be a result of the general limitation on one’s attention or memory resource for the processing of spatiotemporal stimuli. Since both attention and working memory have limited capacity, it is possible that while observing a dynamic object, the visual system sets up a temporal filter for the real-time visual input. Neither the results of the current experiment or the previous work (Chuang et al., 2006; Liu & Cooper, 2003; Spetch et al., 2006; Stone, 1998, 1999; Vuong & Tarr, 2006) could exclude the possibility that the observers divided a view sequence into small chunks and used the statistical representation of each local chunk’s order to represent the whole sequence. If that was the truth, they would not be able to remember a view sequence without consistent local chunk. However, we would like to propose a second interpretation. The presence of local reversal effect in Experiment 2a and the lack of global reversal effect in Experiment 2b might reflect the effect of perceptual organization, which was biased by the relative perceptual saliency of the global and local dynamics in our manipulation. Although in the learning phase, observers were repeatedly exposed to the two objects in fixed scrambled patterns, the lack of saliency of global feature possibly made people unable to remember it or simply do not use it for recognition. On the other hand, the local feature might overwhelm the global feature due to its benefit from the regularity in sequence order (statistically salient). This hypothesis was more consistent with our basic assumption as we predicted the visual system to behave in a more organized way. We assumed that the visual system chose a dynamic feature according to its perceptual properties rather than passively accepted a feature according to its temporal duration.

Experiments 3a and 3b

In the previous two experiments, we had demonstrated the important role of a sequence’s local feature (temporal order within chunks) in recognition. We argued that the visual system used a temporal feature due to its perceptual saliency instead of its time scale. To test this hypothesis, we tried to find out the relationship between feature saliency and its role in recognition in another two experiments. If the global and local features defined by sequence order were really two separate temporal features for the coding of spatiotemporal signature as we assumed, there should be a case in which the global feature could be used. In Experiments 3a and 3b, we expected to find such a condition when the global feature of a sequence was used while the local feature lost its competence by changing the relative saliency of these two features in the sequence.

Methods

Participants

Twenty college students (9 females, average age 22) and 25 college students (12 females, average age 24) volunteered to participate in Experiments 3a and 3b. They were all paid for participation and were naive to the previous experiments.

Design

We followed the design of Experiments 2a and 2b. However, to change the relative saliency of global and local features, we changed the scrambling manipulation to the sequences. Generally, each sequence still contained ten smooth chunks for the perception of apparent motion. However, this time we randomized the rotation direction of each chunk one by one, thus some of the chunks were CW, while the others were CCW (not necessarily half and half). To enhance the global saliency, the displayed order of the ten chunks was NOT randomized as that in the previous experiments (Figure 5). Movie 2 demonstrated an object rotating in the locally scrambled sequence.

Although there are many causes of perceptual saliency, in the present study, we only varied the statistical

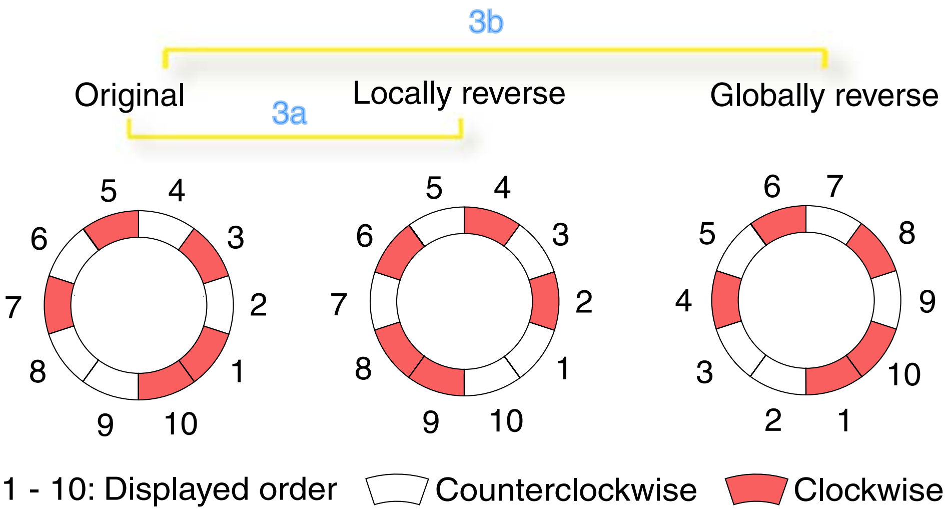


Figure 5. Design of Experiments 3a and 3b. The sequences were scrambled at the local level (inconsistent rotation direction within each chunk), indicated by the colors, but with salient global feature (ordered chunk order), indicated by the numbers. In Experiment 3a, view sequence order of each local chunk (colors) was reversed; in Experiment 3b, the global chunk order (numbers) was reversed.

regularity of the sequence order. This kind of manipulation avoided the distinctiveness caused by some low-level features (such as luminance), which might draw attention automatically. Instead, we expected to observe the effect on object recognition, which was caused by the sequence's temporal properties.

We assumed that the observers would use the global feature (chunk order) for recognition when its saliency was enhanced (keeping in order). By contrast, they might ignore the local feature (temporal order within the chunk) when it lost its statistical saliency (inconsistent among the ten chunks). Thus we predicted that the local reversal effect would disappear when the sequence was locally reversed (Experiment 3a). Instead, we expected to observe a global reversal effect when the global feature of the sequence was reversed (Experiment 3b). Otherwise, if the visual system had a limited-scale temporal window for the spatiotemporal information, we would expect that the non-salient local feature make it hard for the observers to discover the global order underlying the jittering sequences and no global reversal effect would be found in Experiment 3b.

Procedure

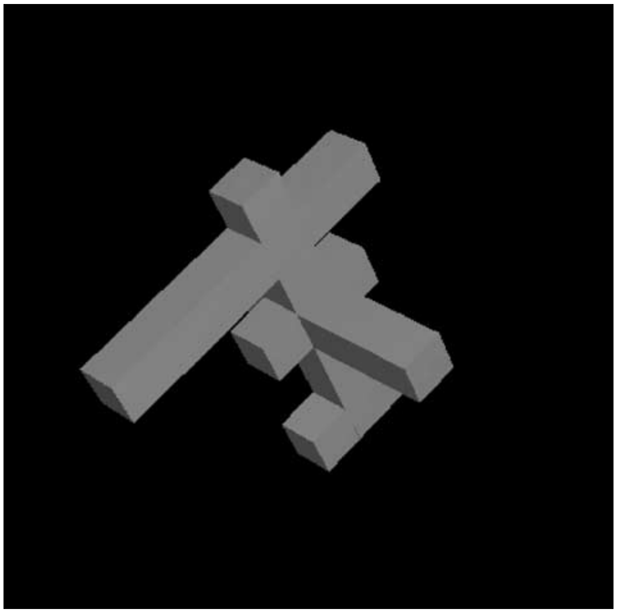
The procedures of Experiments 3a and 3b were exactly the replications of Experiments 2a and 2b, except that we used the locally scrambled sequences.

Results

We refined the data by the same rules that were applied to the previous experiments. One participant failed to pass the criteria in Experiment 3b. The numbers of invalid trials in Experiments 3a and 3b were 48 (2% of total) and 161 (5.59% of total), respectively.

The performance during the learning phase was comparable in these two experiments. Mean accuracies for Experiments 3a and 3b were 94.1% and 93%. Reaction times were 595.5 ms and 600.3 ms, respectively.

Results of the test phase verified our assumptions. By breaking down the regularity of sequence order within local chunks, we found in Experiment 3a that reversing



Movie 2. Demonstration of the stimuli used in Experiments 3a and 3b. The view sequence was locally scrambled (randomized rotation direction within each chunk).

the local dynamic feature had no impairment on discrimination accuracy ($t(19) = 0.647$, $p = 0.263$, one-tailed) and RT ($t(19) = 1.451$, $p = 0.082$, one-tailed). The accuracies for original and locally reverse conditions were 0.95 (*SE* = 0.010) and 0.95 (*SE* = 0.012). RT data were 1708.2 ms (*SE* = 97.9) and 1761.1 ms (*SE* = 106.5). Furthermore, we found in Experiment 3b that reversing the global feature of the view sequences did impair observers' performance in terms of both accuracy ($t(23) = 1.899$, $p = 0.035$, one-tailed) and RT ($t(22) = 3.617$, $p = 0.001$, one-tailed). The accuracy dropped from 93% (*SE* = 0.018) to 82.2% (*SE* = 0.059) from the original condition to the globally reverse condition. The response time increased from 1591.9 ms (*SE* = 75.5) to 1712.5 ms (*SE* = 95).

Discussion

Experiments 3a and 3b provided evidence against the local-advantage assumption that the temporal filter of attention or memory would prevent the observers from using the global order of a sequence for recognition. Thus the lack of global reversal effect in Experiment 2b could not be simply attributed to the general restriction of attention or memory capacity. By contrast, the results supported the assumption that changes in the relative saliency of dynamic features might change the observers' sensitivity to these features, which was consistent with the law of perceptual organization. As shown by Experiment 3b, as long as the global feature was salient enough, it could overwhelm the local feature and be used for recognition. For a similar reason, when the temporal order of local chunks was no longer consistent, it became an inefficient cue for recognition.

General discussion

Decomposable spatiotemporal signature under perceptual organization

In 5 experiments, we have demonstrated that changing motion trajectory to some extent could bias the observers' responses in an object discrimination task. Especially, this effect was selective to the reversal of certain features for specific sequences (Experiments 2a and 3b). The first significant implication of these results was that even artificial spatiotemporal stimuli could be recognized by decomposable features. Despite the controversy about whether or not the recognition of biological stimuli, e.g., point-light walkers, relied on the global form (Beintema & Lappe, 2002; Lange et al., 2006; Reid et al., 2009) or some mid-level optic flow features (Casile & Giese, 2005; Troje, 2002), few studies examined this issue in novel dynamic object recognition. The present study raised the question of whether people coded the view sequence of a dynamic object as it appeared to be or coded the sequence by some kinds of features. The results were in support of the second. We found cases when the observers' recognition performances got impaired when the object's motion trajectory was reversed at either local (Experiment 2a) or global (Experiment 3b) level. In other words, the global (chunk order) and local (temporal order within the chunks) dynamics of an object's view sequence could serve as separate features for recognition. These results shed light on the way we look into the "spatiotemporal signature in mind". First, feature extraction was not a special property for biological motion recognition (Casile & Giese, 2005; Troje, 2002); instead, it might be a general principle for dynamic object recognition. Second, the feature used for recognition was not necessarily image-based as that applied to static object recognition (Ullman, 2007). As in our study, the view images of the objects were the same in all the test trials. The features used for recognition were largely determined by the temporal order of the objects' view sequences, or rather their spatiotemporal appearance.

We argued that the underlying mechanism of feature-based recognition found in the current study might be a cognitive filter, which gated the visual input by its perceptual saliency. In another words, the brain only chose a feature for recognition when the to-be-discriminated motion sequences were salient enough at this specific feature. Notice that spatiotemporal smoothness was not a constraint for this process. The view sequence we employed either had scrambled global temporal order or randomized local temporal order. Moreover, temporal capacity of neither attention nor memory was the bottleneck. As we showed in Experiment 3b, the chaos induced by inconsistent local temporal order did not prevent the participants from extracting the global feature. Previous studies demonstrated that the relative availability of shape and motion might affect the use of spatiotemporal signature (Balas & Sinha, 2009; Vuong & Tarr, 2006). The current study extended the findings by demonstrating that the relative saliency of global and local temporal features of a sequence would also critically affect the coding of spatiotemporal signature. Specifically, we found that the sequences with consistent local order but scrambled global order (Experiments 1, 2a and 2b) and with consistent global order but scrambled local order (Experiments 3a and 3b) were both sufficient to make the local or global feature perceptually adaptable, which implied an automatic spatiotemporal feature coding process under the mechanism of perceptual organization.

Top-down regulation on sequence smoothness in recognition

Although motion smoothness was assumed to be a prerequisite for view association in 3D object learning

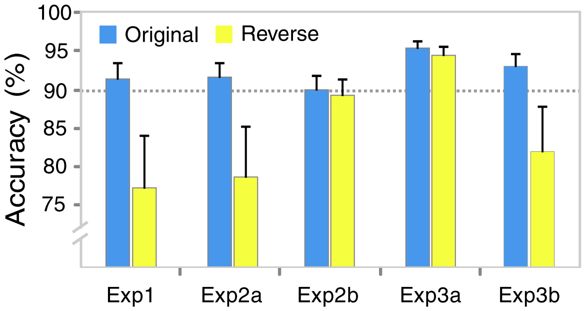


Figure 6. Accuracy of original and reverse conditions in all five experiments. Error bar indicates one *SE*.

(Balas & Sinha, 2008; Wallis & Bülthoff, 2001) and a constraint for optic flow computing (Horn & Schunck, 1981), the present study suggested that the coding of spatiotemporal signature might not be affected by partially scrambling the sequence.

However, before making a conclusion about the role of sequence smoothness on the coding of spatiotemporal signature, we had to test whether the observers could access to identity information from the scrambled sequence effectively. We tried to test this by looking into the relative and absolute difficulties of our experimental manipulations from the recognition performance. Cross-experiment one-way ANOVAs on both accuracy and RT data of the baseline conditions, i.e., the original motion conditions, showed no difference on accuracy ($F(3, 84) = 1.591$, $p = 0.198$) and RT ($F(3, 84) = 1.558$, $p = 0.206$) among the five experiments. These comparisons suggested that the reversal effects we got in Experiments 1, 2a, and 3b were all due to the decrease in reverse conditions (Figure 6), instead of different byproducts under different perceptual difficulties. For the absolute performance, the participants behaved quite accurately in all the baseline conditions (Figure 6), which indicated that the change in sequence smoothness had little impact on their representation of the objects. Post-experiment interviews also found that the observers hardly had any explicit knowledge about what the sequences were really like. Although the sequences we used were always jittering and totally different between Experiments 2a and 2b and Experiments 3a and 3b, the participants in different experiments gave similar descriptions about the sequences. Surprisingly, most of them simply did not notice the incoherence of the objects' motion. More than half of them were not aware of any change in shape or motion during test when being asked. Among the few participants who mentioned the change in view sequence, they just reported the "speed" or "direction of rotation" changed. Taken together, these results implied the coding of spatiotemporal signature was not restricted by the absolute smoothness of the sequence as long as there were available features.

The underlying mechanism of this phenomenon might be an experience- or knowledge-driven expectation. The observers might have a strong tendency to process motion as a statistical property of the dynamic objects for the purpose of recognition (Newell et al., 2004), so they expected to perceive the objects moving in a regular way even when the motion trajectories were artificial and partially scrambled. Previous studies have found the top-down regulation on 3D object recognition when the objects' spatial structures were scrambled (Bülthoff et al., 1998; Sinha & Poggio, 1996). The present study demonstrated the remarkable regulation on sequence's smoothness, which was violated by temporal scrambling. This top-down regulation, combined with the bottom-up organization, reflected the adaptable nature of spatiotemporal coding in dynamic object recognition.

Shape and motion in spatiotemporal signature

We examined the observers' sensitivity to shape change by comparing their RT to the unexpected novel object with that to the studied objects. A constant delay of responses to the novel object than to the studied objects was found among all the 5 experiments (Table 1). Although the participants had no expectation of the novel object and their task was to discriminate the two studied objects, the increase in response time implied that they had detected shape change to some extent, with or without awareness. The robustness of such an ability of shape discrimination together with the reversal effect we found suggested that both spatial and temporal appearances of the objects were coded for recognition. Though the study was not intended to disentangle the facilitating effect of motion on shape processing and the effect of motion as a recognition cue, results of Experiment 1 might imply that there were both in the task. However, we could not draw conclusion from the results that whether the motion information was integrated with the shape (as a spatiotemporal event) or used independently. Although our manipulation on the sequence was based on the temporal order, the consequent change was always spatiotemporal. Actually there is no purely temporal without spatial information in a spatiotemporal signature.

Another potential confound that might need to be addressed was that the observers recognized the objects only by their static views. The recognition performance

	Studied object		Novel object	
	RT (ms)	*SD*	RT (ms)	*SD*
Experiment 1*	1408.42	402.59	1590.48	533.48
Experiment 2a**	1569.21	490.83	1867.63	711.70
Experiment 2b**	1427.31	402.59	1607.65	533.48
Experiment 3a**	1752.86	450.75	2153.90	672.47
Experiment 3b**	1645.30	370.09	1950.71	336.99

Table 1. Reaction times to the learnt objects and the novel objects in all the five experiments; *$p < 0.05$, **$p < 0.01$.

was above chance level even in the reverse motion conditions, which suggested the dominant role of object's shape in recognition. If specific views always appeared at specific positions of the sequence, it would be doubtful whether the observers' responses were biased toward those views. However, the experiment controls on view sequence excluded this confound. First, all the learning trials and test trials contained the same view images (except the novel trajectory condition in Experiment 1). Second, the order of views for each object in the learning and test trials (see Experiment 1, Methods section) was balanced. Specifically, every ten trials of each object started with ten different initial chunks (around the 360 degrees). This Latin square design ensured every specific chunk (so did the views) of the object appeared at a certain temporal position of the sequence with equal probability, thus eliminated the effect of view order.

Implications for future studies

One question beyond the solution of the present study is the generalizability of the temporally global and local features. Though we have observed the dissociation of reversal effects at two temporal scales, the unambiguous definition of global and local features and other potential features that might be used to represent the sequence still need to be clarified in the future. It would be helpful to test different dynamic features within the context of complex and natural motion, such as biological motion. Besides that, we have no definite answer about whether dynamic cues other than motion direction (e.g., speed, curvature) would share a similar working course in the brain with it. For example, is other kind of flow unsmoothness (caused by speed or other flow statistics) tolerable for the coding of spatiotemporal signature?

Conclusions

In the current study, we have examined the coding of spatiotemporal signatures in a dynamic object recognition task. The first important finding is that the mental representation of novel dynamic objects is decomposable. Particularly, the coding process is marked by selective feature extraction under the law of perceptual organization. A second finding is that the way the brain deals with the spatiotemporal signature is quite compulsive. Physical scrambling at a certain temporal scale does not prevent the observers from extracting valid features from the unsmooth sequence.

The results support a framework that highlights the role of spatiotemporal signature in dynamic object recognition. Tolerance of certain spatiotemporal variations makes it possible for humans to recognize objects moving in different manners. At the same time, the automatic organization of view sequence requires some features, such as those defined by global and local temporal orders of the sequence, to be extracted.

Acknowledgments

We would like to thank David Knill and two anonymous reviewers for their helpful comments on the previous version of this manuscript. This research was supported by the National Natural Science Foundation of China, Project Number 30670701.

Commercial relationships: none.
Corresponding author: Dr. Kan Zhang.
Email: zhangk@psych.ac.cn.
Address: 4A Datun Road, Chaoyang District, Beijing 100101, China.

References

Balas, B., & Sinha, P. (2008). Observing object motion induces increased generalization and sensitivity. *Perception, 37,* 1160–1174. [PubMed] [Article]

Balas, B., & Sinha, P. (2009). A speed-dependent inversion effect in dynamic object matching. *Journal of Vision, 9*(2):16, 1–13, http://www.journalofvision.org/content/9/2/16, doi:10.1167/9.2.16. [PubMed] [Article]

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences, 99,* 5661–5663. [PubMed]

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science, 5,* 221–225.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Blake, R., & Lee, S.-H. (2005). The role of temporal structure in human vision. *Behavioral and Cognitive Neuroscience Reviews, 4,* 21–42. [PubMed]

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89,* 60–64. [PubMed]

Bülthoff, I., Bülthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience, 1,* 254–257. [PubMed]

Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision, 5*(4):6, 348–360, http://www.journalofvision.org/content/5/4/6, doi:10.1167/5.4.6. [PubMed] [Article]

Chuang, L. L., Vuong, Q. C., Thornton, I. M., & Bülthoff, H. H. (2006). Recognizing novel deforming objects. *Visual Cognition, 14,* 85–88.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9,* 353–356.

Farivar, R., Blanke O., & Chaudhuri A. (2009). Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *Journal of Neuroscience, 29,* 5336–5342. [PubMed]

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society: Biological Sciences, 269,* 1939–1947. [PubMed] [Article]

Friedman, A., Vuong, Q. C., & Spetch, M. (2010). Facilitation by view combination and coherent motion in dynamic object recognition. *Vision Research, 50,* 202–210. [PubMed]

Friedman, A., Vuong, Q. C., & Spetch, M. L. (2009). View combination in moving objects: The role of motion in discriminating between novel views of similar and distinctive objects by humans and pigeons. *Vision Research, 49,* 594–607. [PubMed]

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). Bold activity during mental rotation and viewpoint-dependent object recognition. *Neuron, 34,* 161–171. [PubMed]

Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Sciences, 97,* 8186–8191. [PubMed] [Article]

Harman, K. L., & Humphrey, G. K. (1999). Encoding 'regular' and 'random' sequences of views of novel three-dimensional objects. *Perception, 28,* 601–615. [PubMed]

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11,* 880–885. [PubMed]

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence, 17,* 185–203.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517. [PubMed] [Article]

Koffka, K. (1999). *Principles of Gestalt psychology*. London: Routledge.

Kourtzi, Z., & Shiffrar, M. (1997). One-shot view invariance in a moving world. *Psychological Science, 8,* 461–466.

Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: A template-matching analysis. *Journal of Vision, 6*(8):6, 836–849, http://www.journalofvision.org/content/6/8/6, doi:10.1167/6.8.6. [PubMed] [Article]

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. *Perception, 23,* 595–614. [PubMed]

Liu, T. (2007). Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory. *Perception, 36,* 1320–1333. [PubMed]

Liu, T., & Cooper, L. A. (2003). Explicit and implicit memory for rotating objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 554–562. [PubMed]

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Biological Sciences, 200,* 269–294. [PubMed]

Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331,* 68–70. [PubMed]

Newell, F. N., Wallraven, C., & Huber, S. (2004). The role of characteristic motion in object categorization. *Journal of Vision, 4*(2):5, 118–129, http://www.journalofvision.org/content/4/2/5, doi:10.1167/4.2.5. [PubMed] [Article]

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6,* 261–266. [PubMed]

Owen, C. B., Tang, A., & Xiao, F. (2003). *ImageTclAR: A blended script and compiled code development system for augmented reality*. Paper presented at the International Workshop on Software Technology for Augmented Reality Systems, Tokyo, Japan.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266. [PubMed] [Article]

Reid, R., Brooks, A., Blair, D., & van der Zwan, R. (2009). Snap! Recognising implicit actions in static point-light displays. *Perception, 38,* 613–616. [PubMed]

Richardson, M., & Johnston, L. (2005). Person recognition from dynamic events: The kinematic specification of individual identity in walking style. *Journal of Nonverbal Behavior, 29,* 25–44.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3,* 1199–1204. [PubMed]

Sarkheil, P., Vuong, Q. C., Bülthoff, H. H., & Noppeney, U. (2008). The integration of higher order form and motion by the human brain. *Neuroimage, 42,* 1529–1536. [PubMed]

Schultz, J., Chuang, L., & Vuong, Q. C. (2008). A dynamic object-processing network: Metric shape discrimination of dynamic objects by activation of occipitotemporal, parietal, and frontal cortices. *Cerebral Cortex, 18,* 1302–1313. [PubMed]

Setti, A., & Newell, F. N. (2009). The effect of body and part-based motion on the recognition of unfamiliar objects. *Visual Cognition, 18,* 456–480.

Siegel, R. M., & Andersen, R. A. (1988). Perception of three-dimensional structure from motion in monkey and man. *Nature, 331,* 259–261. [PubMed]

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature, 384,* 460–463. [PubMed]

Spetch, M., L., & Friedman, A. (2003). Recognizing rotated views of objects: Interpolation versus generalization by humans and pigeons. *Psychonomic Bulletin & Review, 10,* 135–140. [PubMed]

Spetch, M. L., Friedman, A., & Vuong, Q. C. (2006). Dynamic object recognition in pigeons and humans. *Learning & Behavior, 34,* 215–228. [PubMed]

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research, 38,* 947–951. [PubMed]

Stone, J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research, 39,* 4032–4044. [PubMed]

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2,* 55–82.

Tarr, M. J., Bulthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8,* 282–289.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision,* 2(5):2, 371–387, http://www.journalofvision.org/content/2/5/2, doi:10.1167/2.5.2. [PubMed] [Article]

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London, 203,* 405–426. [PubMed]

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11,* 58–64.

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research, 44,* 1717–1730. [PubMed]

Vuong, Q. C., & Tarr, M. J. (2006). Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. *Perception, 35,* 497–510. [PubMed]

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45,* 205–217. [PubMed]

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98,* 4800–4804. [PubMed]