Interpretation of optic flows synchronized with observer's hand movements

Hiroyuki Umemura *, Hiroshi Watanabe

Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology, 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan

A R T I C L E   I N F O

A B S T R A C T

We investigated whether the visual system could use a novel action–perception relationship mediated by a touch panel to resolve ambiguity in 2D optic flows. The stimulus was an optic flow produced by a dotted plane, which was translated and rotated in depth. The translation was synchronized with subject's hand movements on a touch panel. There were two perceptual interpretations of the stimulus as a surface patch oriented in 3D: (1) approaching it in depth and rotating away from gaze normal, or (2) not translating it in depth and rotating toward gaze normal around an axis perpendicular to that of Case 1 [Wexler, M., Lamouret, I., & Droulez, J. (2001a). The stationarity hypothesis: an allocentric criterion in visual perception. *Vision Research, 41*, 3023–3037]. Subjects reported the direction of the axis of rotation, which was perceptually coupled with the perception of translation in depth. The results indicate that the frequency of perception in Case 1 increased as the sessions progressed. This suggests that the visual system learned the association between hand movements and viewpoint translation during the experiment and used this association to decompose the optic flow.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Both the movement of objects and observers' self movements can generate a 2D optic flow pattern on the retina. To reconstruct an accurate 3D structure from optic flows, the visual system needs to decompose a given optic flow into a component generated by object movement and a component generated by self movement. The extra-retinal information derived from motor commands and/or proprioceptive information plays an important role in estimating self movement (Gogel & Tietz, 1973; Wallach, Stanton, & Becker, 1974; Tcheang, Gilson, & Glennerster, 2005). Studies that have compared actively moving observers with immobile observers have revealed that extra-retinal information contributes to the former's precise perception of absolute distances (Panerai, Cornilleau-Pérès, & Droulez, 2002; Peh, Panerai, Droulez, Cornilleau-Pérès, & Cheong, 2002), or to their removal of ambiguities in the extraction of a 3D surface from optic flows (Cornilleau-Pérès & Droulez, 1993; van Damme & van de Grind, 1996; Wexler, Lamouret, & Droulez, 2001).

Wexler and Lamouret et al. (2001) examined the contribution of information from head movement to resolve ambiguity in 2D optic flows. They used ingenious stimuli that could produce ambiguity in 3D interpretations. Appropriate combination of an expansion optic flow, which is produced by translation of a plane in depth (Fig. 1: left), and a scale (compression) optic flow, which is produced by rotation in depth away from gaze normal (Fig. 1: middle), results

in nearly the same optic flow produced by the plane not translating in depth and rotating in depth toward gaze normal, whose axis of rotation is perpendicular to that of the former (Fig. 1: right).[1] When subjects were presented with these stimuli and were asked to report the orientation of the axis of rotation, mobile observers and immobile observers tended to have different interpretations, even though they were given the same retinal images. Wexler et al. found that immobile observers predominantly reported an interpretation that corresponded to the right-hand side of Fig. 1 (after this, we will refer to this axis as horizontal), whereas observers moving toward the stimuli reported a mixture of both interpretations in Fig. 1. One explanation for these results is that moving observers could decompose the flow into expansion and scale components using extra-retinal information, because the expansion component was interpreted as being produced by their own head movements. Moreover, this interpretation can satisfy the stationarity assumption (Wexler, Panerai, Lamouret, & Droulez, 2001), in which an object tends to be stationary in an allocentric reference frame.

We currently use various interface devices, such as joysticks, steering wheels, and computer mice, to move our viewpoint in real or in virtual space. In other words, the actions conducted on these interface devices can externally change our surroundings as our self movement changes. However, it is unclear whether the link be-

---

[1] Note that the stimuli without perspective information on rotation have another ambiguity in the rotational direction around the same axis as shown in the lower part of Fig. 1; however, we were interested in a 90° difference and therefore have not referred to this 180° ambiguity in this paper. We have not mentioned other possible interpretations such as closing or opening a hinged plane, because we presumed rigidity, and no subjects reported such interpretations.

* Corresponding author. Fax: +81 727519961.
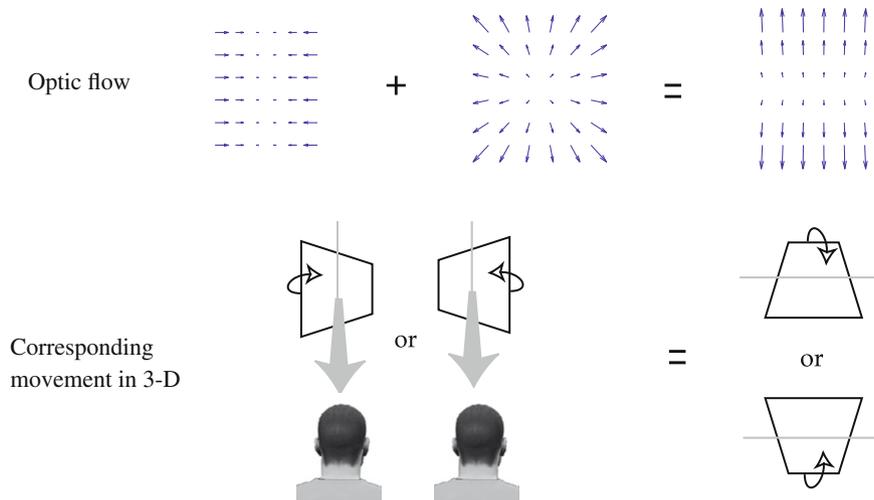E-mail address: h.umemura@aist.go.jp (H. Umemura).

Fig. 1. An ambiguous 2D optic flow produced by rotating plane that can have different 3D interpretations as used by Wexler and Lamouret et al. (2001) and in the present experiment. Appropriate combinations of an expansion optic flow (left) and a scale optic flow (middle) result in nearly the same optic flow produced only by a rotation in depth whose rotation axis is perpendicular to that of the former (right). Corresponding movement in 3D is given in the lower. Because stimuli without perspective information concerning rotation have another ambiguity on rotational direction around the same axis, both these interpretations are given in this figure. We, however, were interested in a 90° difference and therefore did not refer to this ambiguity in this paper.

tween action and perception mediated by these interface devices can have an effect on 3D perceptual processes. The main aim of this study was to examine this problem using a touch panel to translate the viewpoint in simulated 3D space and stimuli nearly identical to those used by Wexler and Panerai et al. (2001). If the extra-retinal information from hand movements on touch panel devices can have an effect on the 3D perceptual process, we can expect to observe the effect of action on the interpretation of 2D optic flows, which have been observed in previous studies.

We assumed three cases for our experiment with regard to the information from hand movements on a touch panel contributing to 3D perception. The first was where the visual system could use information from actions just as it used information from head movements. Here, we expected that subjects would perceive rotation around the vertical axis when they moved the stimulus toward them (their viewpoint approached the stimulus) by their action. This was because the extra-retinal information from their hand movements could be used to decompose the optic flow into an expansion component, which corresponded to translation in

depth, and a scale component, which corresponded to rotation (Fig. 2). The second case was where the information from actions on a touch panel was ineffective to 3D perception. Actions on touch panel would be ineffective if the 3D perceptual processes were only affected by specific action–perception relationships. For example, the action–perception relationship might require extensive experience, or might be established when an action directly translates one's eye position. A mixed interpretation would be obtained in this case of a rotating plane around the horizontal and vertical axes. The third case was where the link between hand movements and translation of the viewpoint was gradually formed during the experiment. Here, we expected that the frequency of perceiving rotation around the vertical axis when subjects moved the stimulus toward them would increase as the experiment progressed. Recent studies have suggested that a short period of visuo-haptic training is adequate to modify 3D visual processing (Atkins, Fiser, & Jacobs, 2001; Atkins, Jacobs, & Knill, 2003; Ernst, Banks, & Bülthoff, 2000; Ernst & Banks, 2002). Furthermore, Adams, Graf, and Ernst (2004) pointed out that the assumption that light
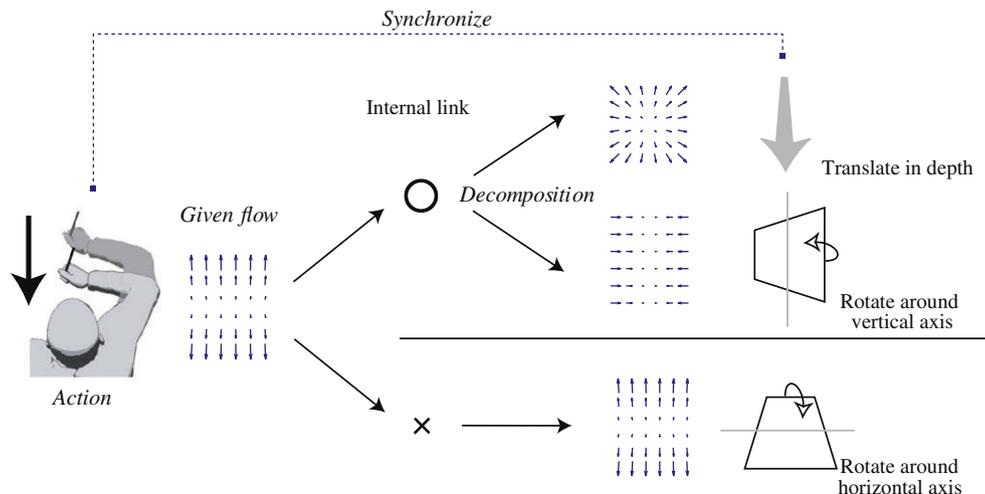

Fig. 2. Two possible interpretations of the optic flow. Upper: a case in which the internal link between hand movement and depth movement are used to decompose the optic flow into expansion and scale components and lower: a case in which the link was not used.

comes from above can be modified after visual haptic training, although the assumption is considered to be hard-wired. We prepared training sessions as these previous studies had done, during which we expected the visual system to establish links between actions and perceptions.

A textured ground, on which the stimulus dotted plane was fixed, was displayed in half the trials in our experiment. The frequency of perception of rotation around the vertical axis should increase in these trials because the presence of the ground plane supports the perception of viewpoint translation in depth. By comparing this "ground-on" condition with a "ground-off" condition, in which the ground is absent, we can observe whether or not the change in the perceived axis during the experiment is purely due to information about an established link between hand movements and viewpoint translation, and not due to other factors such as smoother hand movements acquired during the course of the experiment.

We also examined whether the subjects could establish an association between hand movements and perception and use information about novel action–perception associations to decompose optic flows even when the correspondence between the direction of hand movements and the direction of viewpoint translation was less intuitive. To find out, we divided our subjects into two groups: a forward-and-back and a left-and-right group. The forward-and-back hand movements of the first group were associated with the translation of their viewpoint synchronously in depth, while the left-and-right-hand movement of the second group were associated with the translation of their viewpoint synchronously in depth.

2. Methods

All subjects underwent three types of test sessions and training sessions in the order shown in Fig. 3. In the first to fourth test sessions, subjects moved the stimulus with their hand movements, and were asked to identify the orientation of the axis of rotation. Two other types of test sessions, i.e., an auto-test and a reversed-test session were prepared to confirm our findings. Subjects in the training sessions engaged in a task in which they repeatedly experienced the same novel action–perception relationship as in the test sessions.

2.1. Apparatus

The stimuli were presented on a 21-in. CRT monitor (Sony GDM-F520). The spatial resolution of the monitor was 1600×1200 dots and its refresh rate was 60 Hz. The monitor was positioned 90 cm from the subjects. All viewing was monocular and observers wore an eye patch over their unused eye. The touch panel (Logitec, LTP-17UBK) was 33 cm high and 27 cm wide. Its spatial resolution was 0.63×0.63 mm. Subjects used a pen-sized plastic stylus to touch the panel. We measured the temporal lag

Fig. 3. The order for the experimental sessions.

between the movement on the touch panel and an image displayed by a video camera, and this was 65 ms. The experimental program was developed using OpenGL on a Windows PC. Fig. 4 has an overview of the experimental setting.

2.2. Stimulus

The stimulus was composed of a slanted dotted plane and a textured ground plane in the ground-on condition. The slanted plane was comprised of 100 white dots that were densely distributed near the center and that gradually thinned the further they were away from the center (Fig. 5). This uneven distribution characterized by relatively few dots near the boundaries succeeded in giving the stimulus an irregular shape to prevent the subjects from perceiving its inclination on the basis of dot density and aspect ratio.

We represented the simulated 3D space with a left-handed coordinate system. Each subject's eye position was at its origin and his/her gaze accorded with the z axis and x axis was horizontal. The stimulus was translated on the xz plane by viewpoint translation, which was synchronized with the subjects' hand movements with the stylus on the touch panel. A movement of 1 cm on the touch panel corresponded to a translation of the viewpoint of 5 cm in the simulated 3D space.

The dotted plane was initially placed at $d = 150$ cm, and subjects moved their viewpoint forward-and-back in depth several times before its rotation. Here, d was the simulated depth from the viewpoint to the dotted plane. In this forward-and-back cycle, the d was changed between $90\,\text{cm} < d < 190\,\text{cm}$. The dotted plane was rotated once during approaching movement (=viewpoint proceeding). The rotation started from $150\,\text{cm} < d < 160\,\text{cm}$, and rotated and translated 60 cm in depth (i.e., until reaching $90\,\text{cm} < d < 100\,\text{cm}$). The initial angle of the dotted plane (the plane became frontal when $\theta = 0°$) was $45°$, and became $63°$ after translating 60 cm in depth. The tilt of the dotted plane was randomly determined from $0°$, $30°$, $60°$, $90°$, $120°$, or $150°$ in each trial. This speed of rotation ($0.3°$/cm) in the simulated 3D depth could produce ambiguity in its interpretation shown in Fig. 1. Because the dotted plane was sufficiently small relative to the simulated depth, the dotted plane did not contain perspective information, which should have been produced by the rotation in depth. The average size of the plane at the initial position was approximately $4.0°$.

A ground plane was presented in each of the test sessions with the dotted plane in half the trials (ground-on condition). The ground plane was wrapped with a checkerboard texture and was located 45 cm below the center of the stimulus. The dotted plane was fixed on the plane. The subjects moved the ground plane on the display with the dotted plane by using their hand movements. However, the ground plane did not appear in the ground-off condition.

2.3. Groups

Subjects were assigned to one of two groups, i.e., either a forward-and-back or a left-and-right group. The subjects in the forward-and-back group moved their hands backward and forward and the viewpoint was synchronously translated in depth, while those in the left-and-right group moved their hand right and left and the viewpoint was synchronously translated in depth. Furthermore, we assigned half the forward-and-back group subjects into a subgroup in which tracing upward (away from the subjects) made the camera advance forward in the 3D space and assigned the other half to a condition in which tracing upward made the camera recede. Making the camera recede by tracing upward could be interpreted as subjects dragging the stimulus directly to the far side. The subjects in the left-and-right group were similarly
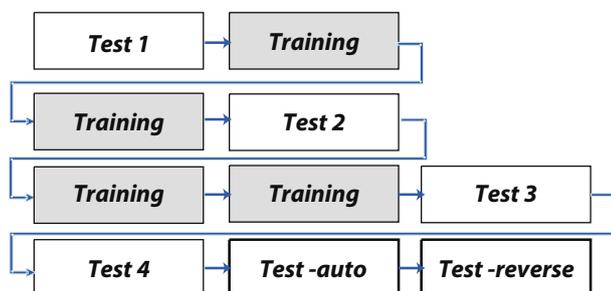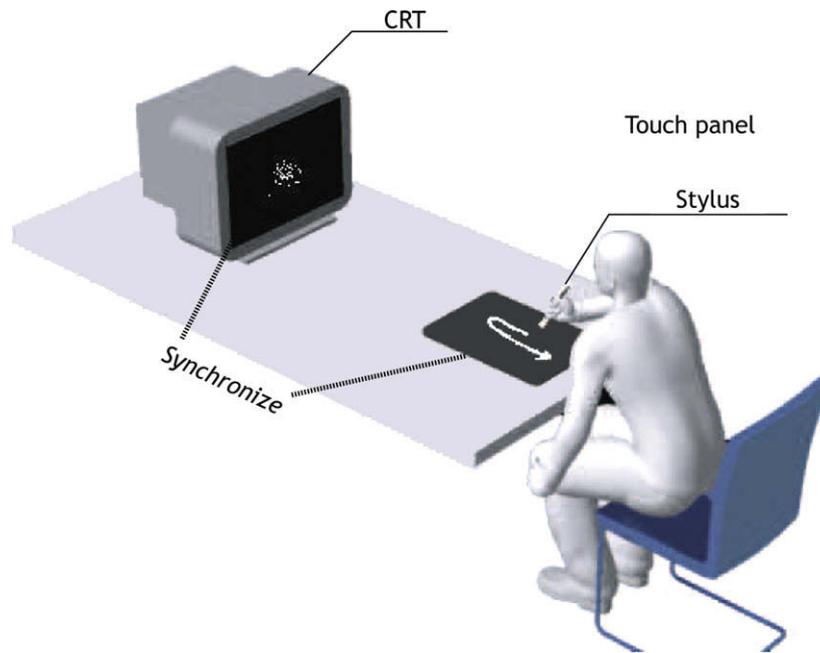
Fig. 4. Experimental apparatus used in the experiment. Subjects monocularly viewed the display. Stimulus moved on the 3D field as a result of viewpoint translation, which was synchronized with the movement of the stylus by the subjects.
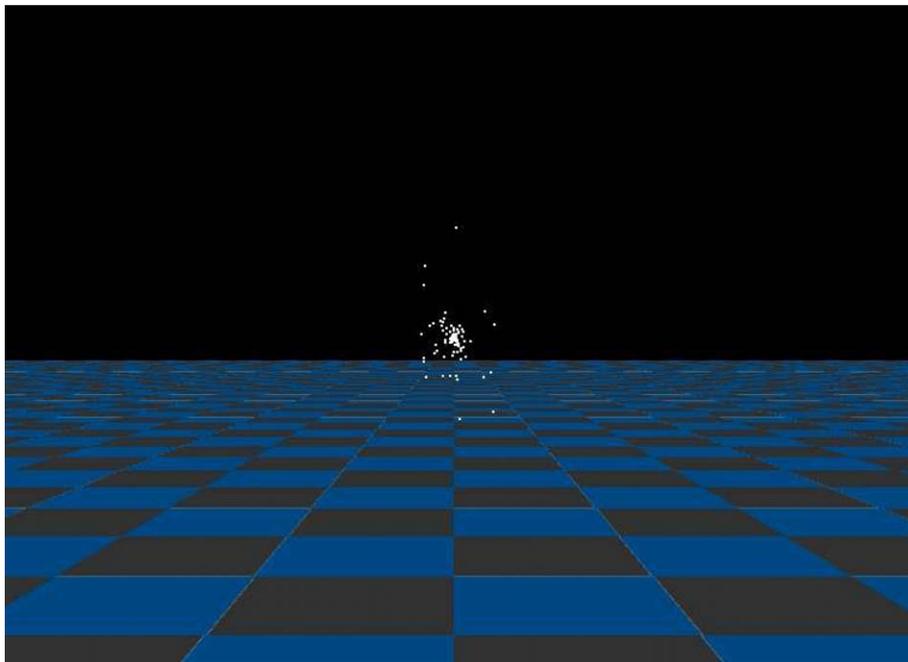

Fig. 5. Clipped image of the center of a stimulus display.

divided and assigned to one of the two subgroups. Hand movement toward the directon of 90° clockwise reltive to the direction of hand movement corresponding to the viewpoint proceeding produced rightward movement of the viewpoint in all the groups.

2.4. Procedure

The experiments were conducted in a dimly lit room. Subjects seated themselves on a chair and placed their heads on a chin rest. The height of the chin rest was adjusted so that the subjects' eye and the center of the stimulus were at the same height.

2.4.1. Test session

In the first to fourth test sessions, each trial began with the appearance of the stimulus on the monitor. The subjects held the stylus just as they would have held a pen and moved it on the touch panel, either back and forward or left and right, depending on the group to which they had been assigned (after this, we will describe the procedure for the subgroup in the forward-and-back group where tracing upward resulted in the camera proceeding.) Subjects were required to keep the point of the stylus on the touch panel during each trial. The viewpoint in each trial was repeatedly translated in virtual 3D space forward-and-back, which was syn-

chronized with forward-and-back hand movements, and rotated once during approaches. The subjects started moving the stylus upward from the center of the touch panel and the viewpoint in 3D space was translated forward in depth. In this viewpoint translation, subjects were required to change the direction of movement to the sound of a beep when the dotted plane reached d = 90 cm and when it reached d = 190 cm. The dotted plane was randomly rotated either in a second, third, fourth or fifth approach in each trial. The stimulus soon disappeared after being rotated 18° in depth. The total duration the stimulus was displayed depended on the timing of rotation and subjects' speed in making hand movements. We required subjects to move the stylus for each forward-and-back cycle within approximately 1 s. Therefore, each display continued for nearly 2–5 s, during which the rotation lasted about 0.3 s.

After the stimuli disappeared, a straight line became visible on the display. Using this line, the subjects reported the orientation of their perceived axis of rotation projected on the frontal plane. They rotated the straight line on a frontal plane every 1° or −1° by pressing buttons. Because the perceived orientation of the axis of rotation and perception of translation in depth co-varied, we were able to know whether the subjects perceived translation in depth indirectly from these responses (perceptual coupling, Hochberg & Peterson (1987)). The dotted plane in this experiment had other ambiguities of 180° with regard to tilt, because they did not contain perspective information on rotation. However, we were only interested in a 90° difference with regard to the interpretation of the axis of rotation, and therefore required the subjects to identify the orientation of the axis instead of the tilt of the plane. The subjects in the preliminary experiment occasionally observed changes in the direction of the axis during a single trial. In such cases, they were required to identify the axis that they had observed for longer durations.

Two questions arose where we could observe an increase in the frequency with which translation was perceived in depth: (1) was hand movement really necessary to perceive translation in depth? (2) was the direction of hand movement essential for perceiving translation in depth? To answer these questions, two other types of test sessions were conducted after the fourth test session. An auto-test session was prepared to address the question about the contribution of hand movements. The stimulus in this session automatically moved along the trajectories that had been recorded in the fourth test session. A reversed-test session was prepared to observe whether the direction of hand movements was essential to perceive translation in depth. As the orientation of the touch panel in this session was reversed, the correspondence between the direction of hand movements and viewpoint translation was also reversed.

Half the trials in each of the test sessions were conducted as part of the ground-on condition, while the rest were conducted as part of the ground-off condition. We also introduced dummy trials, in which the dotted plane was rotated in depth toward gaze normal during translation. Because the direction of rotation in depth was coupled with translation in depth in our stimuli (e.g., when translation in depth was perceived, the direction of rotation in depth was away from gaze normal), we introduced these trials so subjects could not use the direction of rotation in depth as a cue.

As a result, 72 trials (six axis orientations × two ground conditions × five repeats + 12 dummy trials) were conducted in each of the test sessions. The order of presentation in each session was randomized for each subject. Before the experimental sessions, the subjects were instructed to keep their eyes on the dotted plane during each trial, to keep the speed of movement as constant as possible, and to try not to move the stimulus in the x-direction (our stimuli could move in the x-direction). Prior to the first test session, subjects were given 10–20 practice trials to understand

the procedure and to get accustomed to controlling the speed of their hand movements.

2.4.2. Training sessions

The task in the training sessions was to translate the viewpoint in a virtual 3D space to a goal position represented by a dotted plane, the same as that in the test sessions. Settings such as the dotted plane, textured ground, and the correspondence between the movement of the stylus and viewpoint translation were identical to those in the test sessions. A subject's viewpoint in each trial was started from the center of the field. Subjects were required to place the goal on a red square lying on the field at d = 110 cm. This square was fixed on the display. The initial position of the dotted plane was randomly determined in the virtual 3D field, which had a size of 500 × 500 cm in xz space. The dotted plane was occasionally initially positioned outside the viewing volume. When this occurred, subjects were informed in advance that the dotted plane would be visible when they moved their viewpoint backward. The textured field was drawn in half the trials, and but not in the rest. Their order was random. Because absolute information on distance was not included in our experiment, the subjects were required to use size information to adjust the position of the dotted plane. We expected that subjects in this training session could learn the association between stylus movement and viewpoint translation without observing the rotation of the dotted plane. The virtual 3D field used in this training session was so large, relative to the size of the touch panel, that subjects occasionally needed to release the point of the stylus from the touch panel to accomplish the task (this was similar to when one uses a computer mouse with a large display). Subjects, however, were instructed to conduct the task with least number of releases of the point of the stylus from the touch panel as possible. They were also instructed to finish each trial as quickly as possible. The training session ended when subjects completed 100 trials or when their training time exceeded 30 min. On average, the subjects were able to complete the training in about 20–25 min.

Intervals of around 5 min were inserted between each session, while intervals of more than 20 min were inserted after the first training session as well as the second and third test sessions (Fig. 3). All the sessions including intervals took 5–7 h, and were completed within a work day.

2.5. Participants

The forward-and-back group had six subjects (five men and one woman aged between 24 and 38) with normal or corrected-to-normal vision, and the left-and-right group had eight subjects (six men and two women aged between 20 and 36) with normal or corrected-to-normal vision. The participants in the forward-and-back and the left-and-right group were different. With the exception of one participant in the forward-and-back group, none of the participants was aware of the experimental hypothesis.

All the experimental procedures were approved by the Ethics Committee for Human and Animal Research of the National Institute of Advanced Science and Technology. Informed consent was obtained from participants before the experiment.

3. Results

Prior to analysis, we excluded the results of two subjects in the left-and-right group who could not complete half the trials even in the fourth training session within 30 min, although other subjects could complete all the trials within 30 min. These subjects had no systematic responses in the test sessions. We asked subjects to inspect what they saw during the rotation of the dotted plane. Sev-

eral subjects reported that there were several trials in which the direction of the axis changed during rotation, or in which the surface patch seemed twisted. Most of these observations were only obtained in the early sessions. It is worth noting the subjects claimed that they could not distinguish the difference between the two types of rotation: rotation with translation in depth and rotation with no translation in depth.

In analyzing the experiment, the axis responses were represented relative to the given tilt of the dotted plane. We set the axis response to zero when the subjects perceived rotation with translation in depth and correctly reported the perceived axis. For example, when subjects adjusted the probe line vertically for the optic flow in Fig. 1, its axis response was zero. The axis responses ranged from −90 to 90. Since there was no clockwise/counter-clockwise asymmetry, we only considered the absolute values of the responses, which ranged from 0 to 90. We divided the responses into four categories: less than 22°, more than 23° and less than 45°, more than 46° and less than 67°, and more than 68°.

Fig. 6 shows the histograms for the mean probability of subjects' responses in each of the test trials for the forward-and-back group. There seemed to be no salient differences between the two direction subgroups in either of the movement direction groups. Although we could not quantify the difference in the present analysis due to the small number of subjects, two subgroups in each of the forward-and-back group and the left-and-right group were combined. We excluded responses in which the last horizontal position of the dotted plane was largely displaced from the center of the display and in which the last forward-and-back movement took more than 3 s.

We calculated an index, m, for each condition of each observer (Fig. 7) for the ANOVA analyses. The m was calculated with equation $m = p_1 + 0.5 \times p_2 - 0.5 \times p_3 - p_4$, where p_i is the probability of reporting the direction of an axis included in the ith category (p_1; 0–22°, p_2; 23–45°, p_3; 46–67°, p_4; 68–90°). If information from the subjects' hand movements was used to decompose given optic flows into expansion and scale components, m approached one. However, if subjects perceived rotation with no translation in depth, m approached −1.

Before discussing the main results, we should confirm that subjects were able to decompose the optic flows into expansion and scale components when the ground plane gave the impression of translation in depth. The mean m in the ground-on condition was about 0.7 in the first test session in both the forward-and-back and left-and-right groups. This relatively large m in the ground-on condition suggests that subjects would be able to decompose
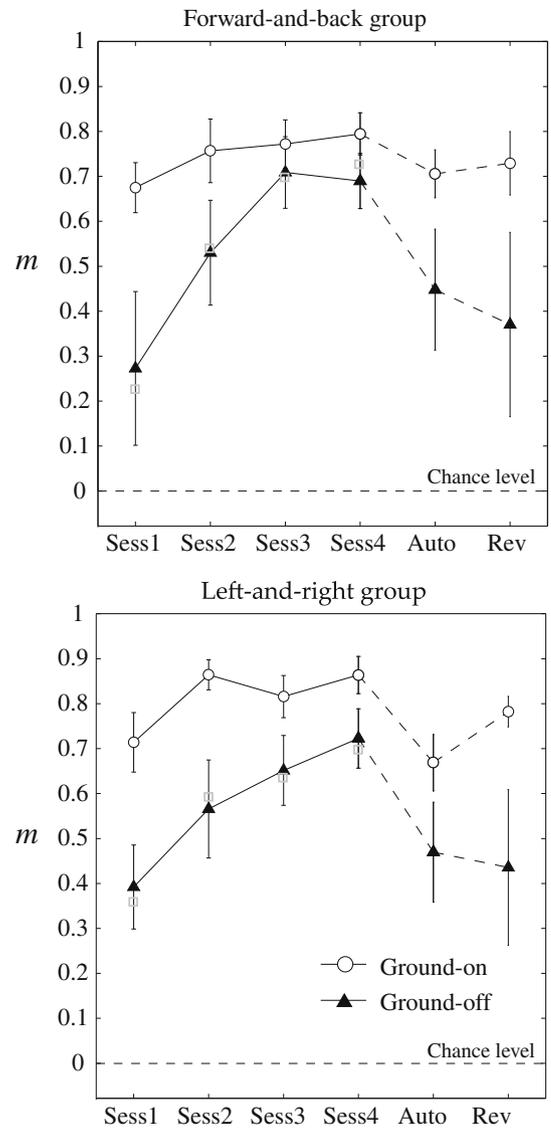

Fig. 7. Mean m in the forward-and-back (upper) and the left-and-right group (lower). Error bars represent between-subject standard errors. In the present task, chance was $m = 0$ and is indicated by the dotted lines. Theoretically, m could reach −1. The open squares indicates mean m derived from the first half of each session.
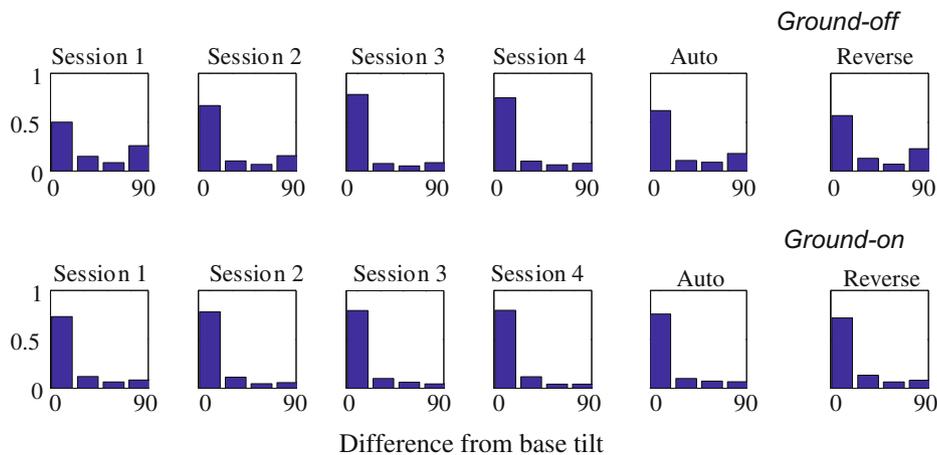

Fig. 6. Mean distributions of responses in test sessions in the forward-and-back group. Upper row represents data from ground-off trials, and lower rows are for ground-on trials.

the optic flow into expansion and scale components by using the accompanying visual information. Warren and Rushton (2008) recently proposed that the visual system uses the relative retinal motion of scene objects to detect movement of an object of interest during self movement. They demonstrated that the global components of motion that the visual system attributes to self movement are subtracted from retinal motion. The presence of the ground plane in our experiment seemed to have this effect on the perception of the rotational plane. That is, we considered that we were able to obtain the same sort of effect as that obtained by head movement as reported by Wexler and Lamouret et al. (2001) in our experimental setting, and could hence discuss the results for ground-off trials in comparison with those for ground-on trials.

To analyze the differences between the four test sessions, we carried out a repeated measures ANOVA with two within-subjects factors: presence of the ground (ground-on vs. ground-off) and session number (first, second, third, and fourth), and a between-subjects factor: movement direction (forward-and-back vs. left-and-right) on m. The analysis revealed significant major effects of the session number, presence of the ground, and the interaction between the two ($F(1, 10) = 15.496$, $p = 0.0001$, $F(3, 30) = 15.668$, $p = 0.0027$, $F(3, 30) = 7.766$, $p = 0.0006$, respectively). In both the ground-off and ground-on conditions, m were significantly changed in the four sessions ($F(3, 60) = 22.691$, $p < 0.0001$, $F(3, 60) = 2.808$, $p = 0.0471$, respectively).

If using information from hand movements became to bias the perception of movement of the dotted plane, we expected that the m in the ground-off condition approached m in the ground-on condition as the session progressed. As expected, a test of the simple major effects of the interaction between the presence of the ground and the session number demonstrated that the effect of the presence of the ground was significant in the first and second sessions ($F(1, 40) = 30.267$, $p < 0.0001$, $F(1, 40) = 15.916$, $p = 0.0003$, respectively); however, this difference did not attain a significance level in the third and fourth sessions ($F(1, 40) = 2.990$, $p = 0.0915$, $F(1, 40) = 3.495$, $p = 0.0689$, respectively). This indicates that the association between hand movements and viewpoint translation were gradually learned, and the information from their hand movements biased perceived movements of the dotted plane as the sessions progressed.

Interestingly, the forward-and-back group and left-and-right group had nearly the same change in m. We could not find significant differences between the two direction of movement groups ($F(1, 10) = 0.259$, $p > 0.5$). The interaction concerning the directions of movement was not significant either between the presence of the ground ($F(1, 10) = 0.088$, $p > 0.5$), between the sessions ($F(3, 30) = 0.445$, $p > 0.5$), or between the three factors ($F(3, 30) = 0.864$, $p = 0.4706$). We anticipated that the mean m of the forward-and-back group would be larger than that of the left-and-right group, because the relationship between action and perception in the forward-and-back group was more intuitive and it seemed easier for them to use cognitive information to interpret the movement of the stimulus; the data indicate that this was not true.

By comparing m in the fourth session and the auto-test session, we could assess whether hand movements contributed to the change in m in the ground-off condition in the test sessions. The difference in m between the fourth test session and the auto-test session was analyzed with a repeated measures ANOVA with the within-subjects factors of the test type (fourth or auto), presence of the ground, and a between-subjects factor of the direction of movement. The m in the auto-test session was significantly decreased from that in the fourth test session ($F(1, 10) = 26.597$, $p = 0.0004$). The effect from the presence of the ground was also significant ($F(1, 10) = 7.627$, $p = 0.0201$); however, the interactions between the presence of the ground

and the test type were not significant ($F(1, 10) = 1.178$, $p = 0.3032$). There were no significant effects from the direction of movement ($F(1, 10) = 0.068$, $p > 0.5$) and the interactions between the direction of movement and the presence of the ground ($F(1, 10) = 0.008$, $p > 0.5$), between the direction of movement and the test type ($F(1, 10) = 0.603$, $p = 0.455$), or between these three factors ($F(1, 10) = 0.238$, $p > 0.5$). From these results, we could discard the possibility that only visual information contributed to the increase in m. The m in the ground-on session also decreased in the auto-test session, which we did not expect. Although the effect seemed small, this implies that the information from hand movements contributed to the perception of the axis of rotation even when the ground plane was present.

If the bias for perceiving translation in depth was specifically coupled with the direction of hand movements in the test and training sessions, we expected that the m would decrease when the orientation of the touch panel was reversed. To find out, we conducted a repeated measures ANOVA with one between-subjects (direction of movement) and two within-subjects factors (test type and presence of the ground) for m in the fourth test session and m in the reversed-test session. The analysis revealed that m was significantly decreased in the reversed-test session ($F(1, 10) = 7.622$, $p = 0.0201$). The analysis also revealed a significant effect from the presence of the ground, and significant interaction between the presence of the ground and the test type ($F(1, 10) = 10.364$, $p = 0.0092$, $F(1, 10) = 5.230$, $p = 0.0452$, respectively). The test of simple major effects on the interaction between the presence of the ground and the test type indicated that the reversal of the orientation of the touch panel only had an effect in the ground-off condition ($F(1, 20) = 12.810$, $p = 0.0019$). These results indicate that the direction of hand movements was associated with the direction of viewpoint translation.

4. Discussion

Our results indicated that the visual system was biased to perceive stimulus translations in depth, which were synchronized with hand movements. This suggests that an association between hand movements and viewpoint translation was established during the experiment and the visual system used this association to decompose optic flows. We think this association between hand movements and translation in the virtual 3D space was mostly learned through the motion of the ground plane in the training session. Because presence of the ground plane would strongly suggests the relationship between hand movements and translation in the virtual 3D space. Moreover, fixation of the dotted plane on the ground could imply that the dotted plane was stationary in the allocentric reference frame in this experimental setting. We think that the change in bias in stimulus perception occurred at a perceptual level rather than at a level of higher cognition, because there were no significant differences between the forward-and-back or left-and-right groups.

Although, we think this association between hand movements and translation in the virtual 3D space was mostly learned through the motion of the ground plane in the training session, our procedure did not exclude the effect the interleaved ground-on condition in the test sessions had on learning the association. In extreme cases, there might be a situation that only this ground-on condition in the test session effectively formed an action–perception association. To find when an association was established, we calculated m with the first half of the trials in each of the sessions. These calculations are indicated in Fig. 7 by the open squares. These m were nearly the same with the m in all the trials, and this indicates that action–perception coupling was largely

established in the test sessions. However, these constant m in each of the sessions suggest that the interleaved ground-on trials contributed to maintaining the effect from learned associations between hand movements and viewpoint translation. As the novel action–perception relationship formed in the present experiment was only effective in our experimental settings, the effect might decrease as the trials in each session progressed if the ground-on trials did not interleave.

In Fig. 7, m was positive even in the first test session and this suggests that there is a certain bias for perceiving depth translation in the first test session. In this regard, we think that the repeating translation of the dotted plane before it was rotated might have increased m through the test session. This preview would promote interpretations of the translated plane in depth in which translation was smooth and continual. However, as revealed by the auto and reversed-test sessions, this preview effect could be separated from the increase in m over the sessions. Another reason for this positive m might be that the interleaved ground-on trials in the test sessions temporally promoted to perceive depth translation in several subsequent trials. However, constant m in the test-sessions indicates that this effect can also be separated from increase in m over the sessions. We should admit that our procedure produced confounding in these points, especially on the effect of the ground. These confounding should be avoided in further researches on the details of this learned association, such as lifetime of the association.

Fig. 8 is a detailed schematic of the process for our results. It is largely based on Kawato, Hayakawa, and Inui (1993), in which the feedforward connection from the lower visual cortical area to the higher visual cortical area provides an approximated inverse model of the imaging process, while the backprojection connection from the higher area to the lower area provides a forward model of the optics. Priors such as rigidity assumptions and object stationarity assumptions have been utilized in the approximated inverse model. As the experiment progressed, particularly in the training sessions, the association between hand movements and viewpoint translation, which can be used to predict optic flows, become more precise and reliable (#1in Fig. 8). This change in association would progress through the use of feedback information derived by comparing a generative image with a given one. Establishing this association between hand movements and viewpoint translation enabled the visual system to decompose optic flows more accurately into the components provided by translation in depth and those by rotation in depth (#2 in Fig. 8).

It has been suggested that visual information is more reliable than haptic or motor information in most spatial tasks (visual capture, Rock & Victor (1964)). However, our results, as well as those from previous studies such as Wexler and Lamouret et al. (2001), have revealed that motor information has biased visual perception. This would be because visual information failed to provide a unique solution against given optic flows in these experiments, and in such cases, we can observe perception biased by non-visual information. Landy, Maloney, Johnston, and Young (1995) proposed that different depth cues provide qualitatively different information, and these qualitative differences must be taken into account by combination rules. Our results can be considered as an example of "promotion" between visual information and extraretinal information, in which depth information acquired from hand movements complements ambiguous visual depth information.

Haijiang, Saunders, Stone, and Backus (2006) recently demonstrated that the visual system can be conditioned to use new visual cues, which are essentially irrelevant to the appearance of a stimulus before experiments, during the perception of a bistable stimulus. The subjects in their experiment could associate the appearance of a stimulus with its position or its translation. One can interpret these results as an instance of cue recruitment, in which hand movement was learned as a cue that affected the perceptual interpretation of the stimulus. Additional work might be needed before making a connection with cue recruitment, especially if subjects participate in experiments with a bias to expect translation in depth during hand motion.

5. Summary

The present study demonstrated that the human visual system can use novel action–perception relationships mediated by a touch panel to resolve the ambiguity in 2D optic flows. This internal use of information from hand movements was similar to the use of information from head movements observed in earlier studies. The findings from our experiment would help in better understanding how perception and action interact to enable accurate and rapid perception. However, we need more studies on lifetimes of the effects of learning and adaptability of the effect to different settings. Furthermore, we need to know about other types of actions through interface devices that may have an effect on 3D perceptual processes.
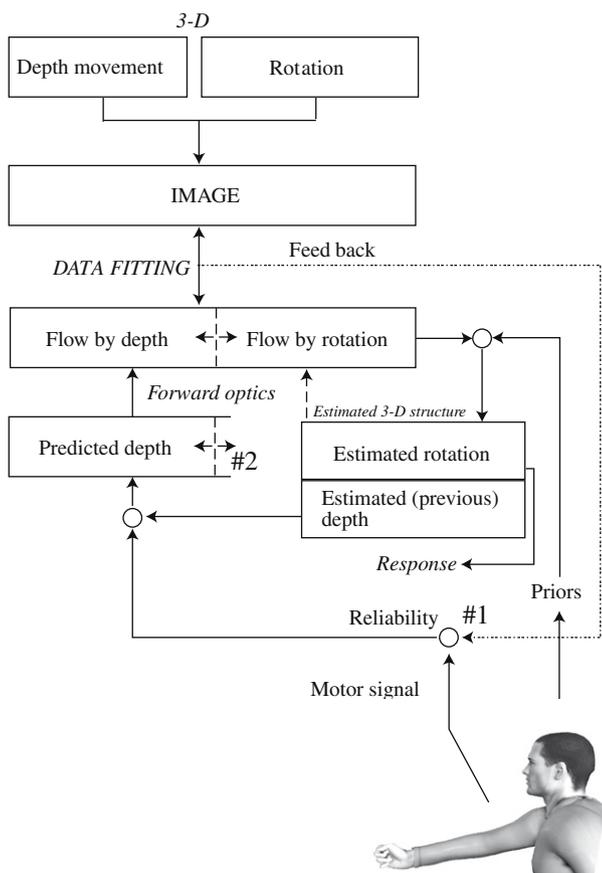
Acknowledgments

Fig. 8. Schematic of contribution of extra-retinal information on estimating the direction of the axis of rotation in the present experiment.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2009.02.020.

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the'light-from-above' prior. *Nature Neuroscience, 7*(10), 1057–1058.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41*(4), 449–461.

Atkins, J. E., Jacobs, R. A., & Knill, D. C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research, 43*(25), 2603–2613.

Cornilleau-Pérès, V., & Droulez, J. (1993). Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structures. *Perception and Psychophysics, 54*(2), 223–239.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433.

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3*(1), 69–73.

Gogel, W. C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception and Psychophysics, 13*(2), 284–292.

Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 483–488.

Hochberg, J., & Peterson, M. A. (1987). Piecemeal organization and cognitive components in object perception: Perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General, 116*(4), 370–380.

Kawato, M., Hayakawa, H., & Inui, T. (1993). A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network: Computation in Neural Systems, 4*, 415–422.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research, 35*(3), 389–412.

Panerai, F., Cornilleau-Pérès, V., & Droulez, J. (2002). Contribution of extraretinal signals to the scaling of object distance during self-motion. *Perception and Psychophysics, 64*(5), 717–731.

Peh, C., Panerai, F., Droulez, J., Cornilleau-Pérès, V., & Cheong, L. (2002). Absolute distance perception during in-depth head movement: Calibrating optic flow with extra-retinal information. *Vision Research, 42*(16), 1991–2003.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science, 143*, 594–596.

Tcheang, L., Gilson, S. J., & Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research, 45*(16), 2177–2189.

van Damme, W. J., & van de Grind, W. A. (1996). Non-visual information in structure-from-motion. *Vision Research, 36*(19), 3119–3127.

Wallach, H., Stanton, L., & Becker, D. (1974). The compensation for movement-produced changes of object orientation. *Perception and Psychophysics, 15*(2), 343–399.

Warren, P. A., & Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research, 48*(5), 655–663.

Wexler, M., Lamouret, I., & Droulez, J. (2001). The stationarity hypothesis: an allocentric criterion in visual perception. *Vision Research, 41*(23), 3023–3037.

Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature, 409*(6816), 85–88.