Eyes Draw Auxiliary Lines in Interpreting Images

Yukio Ohsawa
Department of Systems Innovation, School of Engineering, The University of Tokyo, Tokyo, Japan
Email: ohsawa@sys.t.u-tokyo.ac.jp

Yusuke Maeda
School of Engineering, The University of Tokyo, Tokyo, Japan

***Abstract*— This study aims at finding a typical pattern in the eye movement, when human obtains an interpretation of a presented image. We conducted experiments where subjects looked at abstract artworks and commercial posters. We compared the eye track lines with subjects' statements about the interpretation of the image. The results show the tendency that eyes draw lines in the velocity between saccade and pursuit eye-movement as we call *slow saccade line* (SSL), following objects in the image corresponding to the meaningful units (sentences) of subjects' statement about image interpretation. Furthermore, when they reached an insight, i.e., obtained a new interpretation after feeling difficulty in understanding the image, the eye track changed into SSLs after exploring in the velocity of fast saccade. We found this shift of velocity enables to acquire long-term memory of the obtained interpretation.**

***Index Terms*— Auxiliary lines, slow saccade line, image interpretation, insight**

I. INTRODUCTION

The movement of eyes has been shown to provide a measure of attention in human's decision making, as in consumers' brand choice in the market of products [1, 2, 3]. For example, when a customer of a supermarket is looking at a shelf, her view may be focused on the commodity she has been latently looking for. By investigating the eye movement, we may be able to investigate the latent interest of customers.

There have been significant advances in the modeling of eye-movement, for the recent years. The basic models have been developed based on the time spent for fixating on an object [4, 5]. Methods for automating the protocol analysis of eyes have also been developed based on the cognitive models fixation and movement of eyes [6].

Cognitive scientists linked eye movements to high level (semantic) information processing, e.g., Terai and Miwa discovered that, the direction of eyes' movement changes discontinuously when human obtains insight, i.e., gets released from impasse, i.e., previous constraints on thought in problem solving [7]. For example, in Fig. 1, we find some digits in the cells of the presented table. Terai and Miwa made a system where the lines scroll up slowly, while the subject with eye-tracking glasses looks at the cells. The subjects have been instructed to look at the table to search a rule for computing the values in the third column, e.g., "X + Y = Z," and to utter what he is thinking all the time [7].

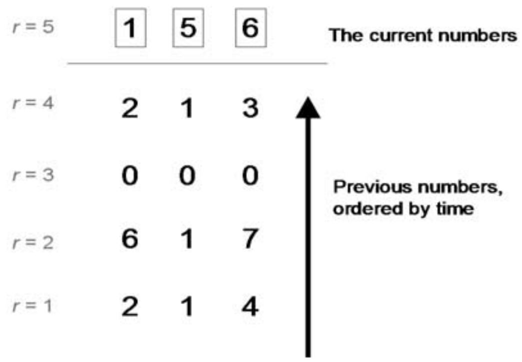

Figure 1. The experiment of finding the rule underlying digits (Terai and Miwa 2003 [7]). The digits for X, Y, and Z (the digits in the right, the central, and the left cells) in each line appear simultaneously, and new lines appear from the bottom.

In the case of Fig. 1, insight meant to be released from the blocking hypothesis "X + Y = Z" based on which subjects first tried to interpret the rule underlying the first three lines. As a result, it has been discovered that, when an insight occurs, the eyes' movement changes. In the case of Fig. 1, the movement was in the horizontal direction when a subject was saying "maybe X+Y=Z, but the forth line violates this rule…" However, when the subject found the correct rule, i.e., "$X_r = X_{r+1} - 3$" for all *r*, the eyes suddenly began to move in the vertical direction. From this example, we can expect the eye-movement may provide a measure of *insight* (see also [8]), that is an essential cognitive effect for understanding the structure of the target problem.

In this paper, we deal with the problem how a viewer interprets a scenario meant by an abstract image. For example, Picasso meant the scenario "the air force attacked innocent people in the city, people and animals are dead on the road, and women are crying hugging babies dead in their arms" by his picture Guernica. The aim of our study means to investigate human's cognitive mechanism in obtaining an interpretation, which is a hypothetical scenario, by looking at an image for a short time. That is, one creates a hypothesis of a scenario underlying an image, for interpreting an image, whether or not the hypothesis corresponds to the scenario meant by the person who created the original image. This can be regarded as an example of the problem how human suddenly interprets an image which had been designed as a medium of a message.

In this paper, we had subjects look at images for a fixed length of time, and investigated the fragments of message the subjects perceived from the image. To summarize the results in advance, a *slow saccade line* (SSL), a continuous line of eye movement as slow as pursuit but occurring in gazing at a still image, corresponds with the unit of viewer's interpretation of the image, according to the experiments executed here.

II. Eyes, attention, and interpretation

A. *Hypothesis : Auxiliary lines emerge in eye tracks in interpreting an image*

In the domain of decision support information systems, *chance discovery* has been studied as a topic for the development of methods to discover a *chance*, an event which may be significant for human's decision making. Because a chance might be rare and its meaning is too uncertain to interpret from the observable part of the real world, data-based computational analysis for understanding a chance is difficult. In this sense, the cognitive process of human for interpreting and understanding the meaning of a chance can be regarded as a crucial issue in the studies on chance discovery.

Researchers on chance discovery have been studying the perception of decision makers in interpreting a map of the co-occurrence relations among events in the market. This map has been obtained by visualizing the co-occurrence of items which represent events recorded in the data from the business environment (e.g., [9,10,11]). The model commonly supported by these researchers and business users of map visualization tools is that a user follows the four steps below, when they discover significant chances from the map:

Step 1) Human first percieves the existence of patterns of event-occurrence which are easy to understand.

Step 2) He explores new scenarios by connecting events in the easy patterns percieved in Step 1.

Step 3) He pays attention to *bridges*, i.e., events which may be rare but relevant to multiple patterns in Step 1.

Step 4) He may (say "aha!" and) obtain a scenario as a guideline for choosing actions, from combinations of patterns in Step 1, via bridges in Step 3. After this moment, he shifts attention to looking around widely over the map.

In step 1) here, the human can be captured by impasse due to cognitive constraints which force him to consider only easy scenarios corresponding to patterns in Step 1. In Steps 2) though 4), he comes to be released from the constraint and creates new scenarios. We expect in this paper that this process model stands also for human's daily interpretation of (imagination of scenarios from) images, not only for a puzzling target as a map of events in the market as in [11] or the scrolling table in Fig.1, but also artworks and commercial posters.

For steps 2) and 3), human should find auxiliary lines [12, 13] between basic patterns, to acquire effective clues to create scenarios.

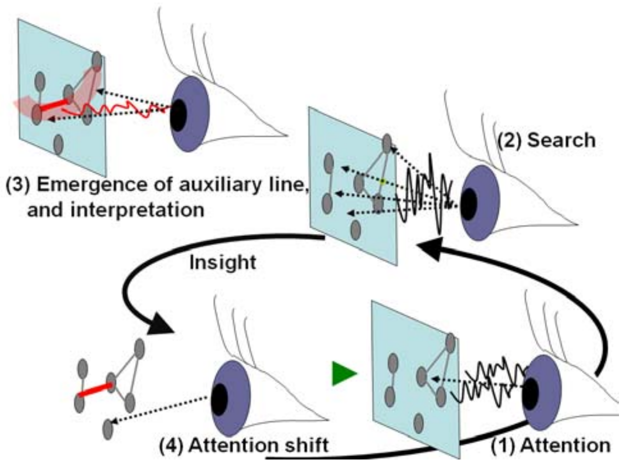

Figure 2. Human's cognitive process for attention and interpretation of an abstract image

We aim in this paper to show the eye movement reflects the latent feeling of viewer underlying the auxiliary lines, which corresponds to the viewer's interpretation whether or not the interpretation is uttered or wrote verbally. From this aspect, we hypothetically extend the process model above as:

Step1') *Initial attention*: The viewer first pays attention to basic and easy-to-understand relations among items in the image.

Step2') *Search*: Then, the viewer explores new scenarios by connecting the basic items in Step 1', sometimes via other parts difficult to understand. The eyes move quickly and widely, as saccade.

Step3') *The emergence of auxiliary lines*: The viewer creates an interesting scenario by sensing/creating a line to connect multiple easy parts. Here, we the eyes move slower than in Step 2), because the viewer thinks logically about the scenario to be acquired .

Step4') *Attention shift*: The viewer changes attention to other parts of the target image.

If this hypothesis stands, we can extract the viewer's unobservable interpretation of the image from the eye-track data, as a slow linear movement of eyes. This will provide a method for evaluating the effect of a commercial poster, on which the viewer is expected to reach an understanding concrete enough to buy products, and of an abstract artwork, which the viewer may interpret as a message of the painter.

III. Experiment (i) eye movements over Abstract Pictures

In this section, we present the method and the result of experiment where we took 10 subjects instructed to look at images which look are new to the subjects - four abstract pictures painted by Pablo Picasso.

A. *The Method of Experiment*

As abstract pictures of which the underlying meanings are said to be hard to interpret but possible to explain once the meaning is interpreted, we showed photographs of pictures by Pablo Picasso to different 5-subjectgroups. For example. the picture in Fig. 3 is the photograph of the central part of the large picture *Guernica*, famous for being an artwork on warfare, but the subjects could not explain the meaning (e.g., "*a scene of air-force attack*") of this figure, according to our preliminary interview.

The pictures were put at the distance of 1.5m from the subject, and the subject wore glasses as in Fig. 4, the eye movement sensing system FreeView produced by Takei Scientific Instruments Co. Ltd in 2002. The eyes-track data were obtained as far as the view angle is within 20 deg from the center of the sight, by the sampling rate of one per 0.033 sec. The relative movement of the eyes to the target has been obtained and drawn in a complex curve as in the left-hand side of Fig. 4. We put target images in the range of 20 deg by putting the image within 0.5m height and 1.5 m front of the subject's eyes.

B. Results

First, let us pick one subject. The eye movement for all the 3 minutes while the subject was looking at Fig. 3 was as in Fig. 5. For details of the changes in the movement, let us see Fig. 6. The number above each picture means the time steps after the beginning of the subject's looking at the image, where 30 corresponds to 1 second. For example, "2500-3500" above Fig. 6-(a) means that the dots in this figure represents the eye movement from 83.3 seconds though 116.7 seconds after the view started.

In the case of this subject, The eye movement started from two parts of Guernica, i.e., the face of the horse in the left of the picture, and the two people in the right. This attention lasted for about 2 minutes.

Next, a thin line from the horse face appeared as in Fig. 6-(b), in the direction to the people in the right. This new line was then extended and emphasized, the attention of the eyes was paid to the line connecting the right and the left parts of the picture, as in Fig. 6-(c).

In Fig. 6-(c), we find a drastic change in the movement of eyes, especially for the 6.7 seconds just before the subject's insightful comment "The light in the hand of the citizen sheds might be a bomb… killing this horse and soldiers! The two citizens in the right-hand side needed the lamp because the bomb attacked and darkened the city." That is, he found the lines (by which Picasso meant light rays) from the bomb and from the lamp connect the right and the left of the picture, and form a meaningful structure in the picture.

This sequence means the subject first paid attention to the easiest parts to interpret, i.e., the eyes of the horse and of citizens, and then searched and found the bridge to connect these basic parts. Thus, steps 1, 2, and 3 in the process model we presented in Section II stands in this case. That is, the emergent interpretation of the figure, represented by the eye movement for the 6.7 seconds just before insight, corresponds to the interpretation of the picture, i.e., as a picture of air force attack in war.

A similar phenomenon occurred for the picture *A Lady with a Hat* by Pablo Picasso as in Fig. 7. Just after the subject's eyes drew lines for 10 seconds on the part connecting the nose-like part and the part like the end of the lady's hair, the subject said "this is a woman, and very attractive!".

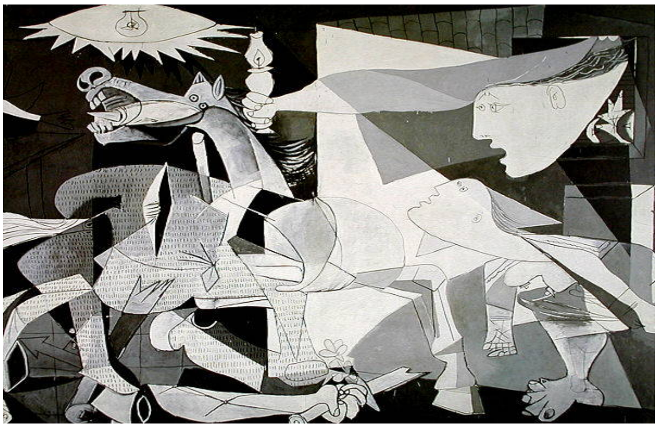

Figure 3. The center-part of *Guernica* by Pablo Picasso

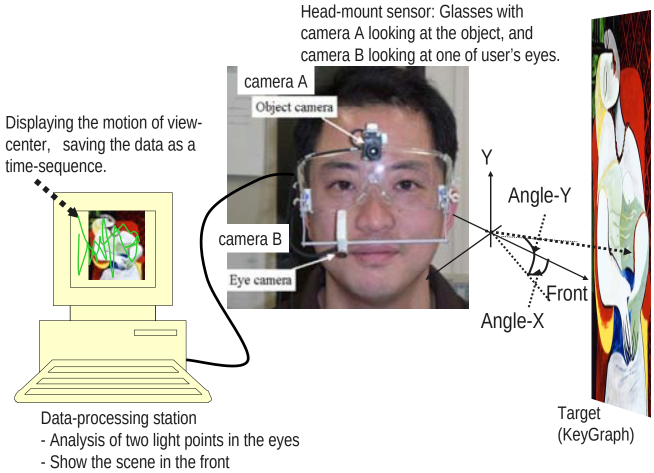

Figure 4. The sensing system (glasses) of eye movement

For all the 20 cases, i.e., 5 cases for each of the 4 pictures presented, the tendencies was significant. That is, 12 of the 20 cases reached interpretation of the picture within the three minutes they looked at the picture, and the lines of eye tracks in the velocity of 10 - 40 [deg/sec], for between 6 and 15 seconds, were found in 8 of the 12 cases. Just after this, they spoke words about their interpretations, corresponding to the lines of eye movement. On the other hand, for the 8 other cases where the subjects did not clearly mention about their interpretation, such lines were found for only 2 cases.

Figure 5. The eyes movement while looking at Figure 3.

6-(a)

6-(b)

6-(c)

Figure 6. The eyes movement for ten seconds before the insight statements

C. Discussion

In [14], Norton showed *scan paths*, lines connecting major components of a picture, play essential roles in human's cognition of the artwork. On the other hand, the result of experiment in this section implies that some short part of the scan path can be regarded as the auxiliary line for combining components of the target image, to achieve a jump to the perception of a meaningful structure of the target image.

Figure 7. The eye movement for 10 seconds just before the subject's insightful comment, for *A Lady with a Hat* by Picasso.

We should point out the speed of eye movements on these lines were not as quick as average saccade which has been regarded as typical movement of eyes in tracking moving objects and exploring things in the real space. Expecting this specific velocity may be useful for detecting essential auxiliary lines from human's eye track data, let us present our following experiments in the next section.

IV. Slow Saccade Lines in The Eye-tracks

Eyes take different types of movements in fixing attention to a still object as:

(Slow) drift: A slow movement taking place invariantly, of about 5 minutes of view angle

Tremor: A high-frequency small tremble of about 1 minute swing which tends to co-occur with drifts

Micro-saccade, or *flick*: A stepwise movement ranging up to 20 minutes which recently is said to be linked to hidden consciousness [15].

And, when the target of attention moves, the movements of eyes are classified on the type of viewer's attention, and the types correspond to different velocities of eye movement as:

Saccade (50 or 60 - 600 deg/sec): A large swing which occurs when the viewer shifts the target of attention. Intentional shifts of attention are reflected to saccade.

Pursuit (5 - 30 deg/sec): Middle-velocity movement occurring in looking at a moving object

Our finding in the last section was novel from the aspect of eye-movement velocity. The eyes move by 10-40 deg/sec, i.e., far slower than average saccade, and a little faster than the range of pursuit (5-30 deg/sec), although the images were still. Although some quick movements, as fast as saccade, have been observed in the experiments, the utterances of the subjects (viewers) about their impressions matched the tracks of these pursuit-like lines.

In the remainder of the present paper, we call such a line a *slow saccade line* (*SSL*) and aim at validating our hypothesis that SSL reflects human's interpretation of target image.

V. EXPERIMENTS ON THE ROLES OF SLOW SACCADE LINES

As experimental targets, here we took commercial posters for product advertisement. To mention about the result outline here, the experiments show that the unit of semantic interpretation of an image corresponds by high accuracy with an SSL. And, the interpretation tends to be recollected after the experiment, if the interpretation matches with the set of objects on one SSL.

A. Eye-tracking Experiment on the corredspondence of SSLs and Uttered Interpretations

Here, we extracted each SSL as a sequential eye-track line, in which eyes move continuously and as slowly as pursuit, but which is not continuously linked to other sequences via similar movements. More specifically, an SSL is taken as in the following algorithm: For each moment t_1 in the period of viewer's looking at the target image, take the longest period from t_1 satisfying two conditions: *Cndt* 1: The eyes move slower than 60 deg/sec, *Cndt* 2: The movement direction does not change by more than 45 deg. If this period is longer than 0.2 sec, take the line as an SSL.

[The algorithm for taking an SSL from eye track data]

Take each time t_1 from 1 through 300 in the data of eye track, where t_1 corresponds to 0.033 t_1 [sec] from the start of experiment for one subject (the 300 time points mean 10 seconds). For each t_1, do :

Take the longest period [t_1, t_2] where each *loc*(τ), the eye's focus at time τ in [t_1, t_2], satisfies both conditions below:

Cndt 1: *loc*(τ) is within a fixed distance δ (set to 2.0 deg of viewer's eye angle) from *loc*(τ-1) .

Cndt 2: The arrow directed from *loc*(τ-1) to *loc*(τ) is within a given angle difference ε (set to 45 deg here) from the arrow from *loc*(τ-2) to *loc*(τ-1), if the distance from *loc*(τ-1) to *loc*(τ) is longer than 0.3.

If t_2-t_1 is larger than 6, take the sequence of *loc*(τ) for all time τ in [t_1, t_2], as an SSL.

Note that 2.0 deg per 0.033 sec means 60 deg per sec, which is twice as fast as the previously said upper limit speed of pursuit and the lower bound speed of saccade. That is, an SSL is an eye track line where the eyes move at a speed close to pursuit on a still image, on which the eyes do not turn the direction suddenly. In Fig. 8, for example, the dots are the eye focus positions for the 10 seconds of one subject. Only the dot sequences in the small square frames in Fig. 9 are the SSLs.

The dots in the square frames A, B, and C in Fig. 8 are magnified in Fig. 9. The five dots appearing sequentially in the eye track of the viewer is in a coherent direction, i.e., always within 45 deg from the movement of just before, and within 1 deg of view angle from the focus position of one step before. On the other hand, the dots (eye focuses) in area B does not have enough number of neighboring dots to form an SSL. And, area C has a dense cluster of dots, but the direction of movements changes frequently by more than 45 deg. Thus, area A has an SSL, whereas areas B and C do not.

The method of experiment

In this experiment, we hired 21 subjects s_i (i=1, 2, …, 21), using the eye-tracking system, and put the target image in front of each subject sitting still. Each subject looked at images of commercial posters and in sequence, each image for a short time. We also conducted free-answer questionnaires about their memory of how they interpreted the presented images, and thus obtained the data of eye tracks and on interpretations of images. Let us describe the details of these data below:

Data 1) The eye-tracks: We used the eye-tracking system as we used in Section III. We set up the same experimental condition, i.e., the sampling rate of one per 0.033 sec, putting target images within 0.5m height and at 1.5 m front of the subject's eyes.

Each subject looked at the images of 10 advertisement posters in Table I presented in sequence, each image for 10 seconds. While looking at each picture, the tracks of eye-movement were recorded in data *loc*(t), where t is an integer value meaning the time counted discretely from the start of the 10 seconds for one image, i.e., 0.033 t [sec] from the start.

Data 2) Questionnaires: Free-answer questionnaire was executed, after looking at images. For reducing the time from viewing each poster to the questionnaire about the poster, we separated the 10 posters for one subject into two cycles. In each cycle, we showed 5 posters and then immediately executed the questionnaire asking "write freely about the posters you can recollect, in the order you recollect easily." All subjects answered this. Here, we chose not to ask after each poster, in order to reduce the risk of eye-movements' bias due to the frequent involvement in the questionnaire.

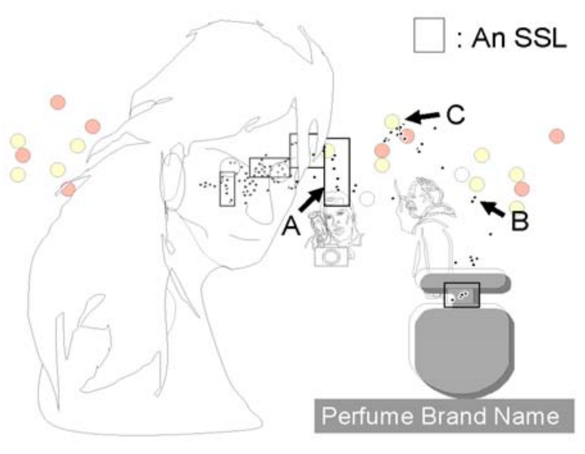

Figure 8. Right: Eye track plotted, and drift lines extracted (subject s_1's eye track in looking at a poster image)

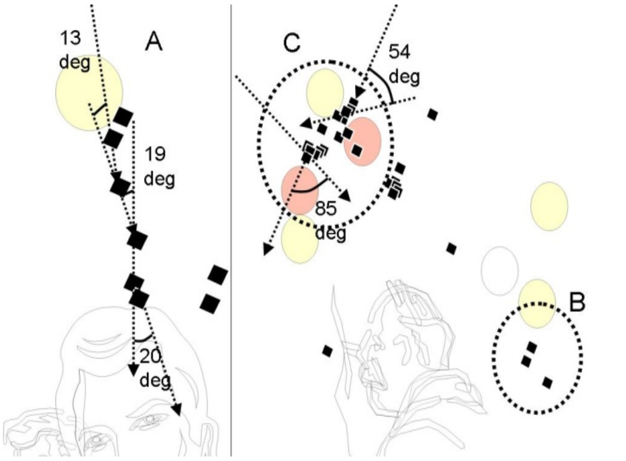

Figure 9. The eye tracks for Figure 8: Area A in this figure has an SSL, whereas B and C does not.

TABLE I.
THE IMAGES USED FOR THE EXPERIMENTS

Image	*The poster's content*
Poster 1) A perfume and an actress	An actress in the left, with a bottle of perfume. People with cameras are behind.
Poster 2) A vinegar for sushi	Ten pieces of sushi, put upside down, with a very small bottle of sushi vinegar
Poster 3) A company of genomics	Ants under the earth surface, and a few sentences about the health of the cells of human
Poster 4) A scientific museum	Boy has a toy rocket, a scientific museum's product
Poster 5) A campaign to help the poverty from hunger	A loaf of bread, shaped like the earth. Letters "bread for all over the world" on the loaf.
Questionnaire	
Poster 6) is Poster 1) again	
Poster 7) A watch for women I	A famous supermodel, having a watch, and a magnified picture of the watch.
Poster 8) A ball point pen	A woman, in close with red and black textiles, is lifting a pen with red and black parts.
Poster 9) A watch for women II	An actress wearing jewelries in a café, with a magnified picture of a watch in the right.
Poster 10) A watch for men III	An magnified face of an actress, showing her arm wearing two man's watches.
Questionnaire	

The method of analysis

From the answers by all the 21 subjects, we computed the rate of words, sentences, and comments matching some SSL in the eye track of each subject. More specifically, we computed the correspondences (SSL-likeliness) in the following manner.

In the discussions below, let us take $Q_{h,i}$, the set of words included in the h-th unit of word sequence in the answer of subject s_i to the questionnaire. A unit here can be one *word*, one *sentence*, or one *comment* (full comment about one image). Here, we define a *word* as a word or a phrase, which corresponds to an entity in the subject's *comment*, which means the written answer to the questionnaire about one image. For example, "a woman," and "a perfume bottle," are both called a word, but a "woman with blond hair buying a perfume bottle" includes three words because three entities (woman, hair, and bottle) are included. We define a *sentence* as a shortest sequence of words expressing an impression of the scene the subject saw, which is independent of the words before and after if any. If the subject inserted apparent delimiters such as periods or a line change, we used them as the ends of sentences. For all answers, three co-conductors of the experiment checked if the obtained sentence satisfies the definition above of a sentence.

Thus, $Q_{h,i}$ was obtained for each pair of h and i. We evaluated the degree of *correspondence Cr*, between an SSL for the time range $[t_1, t_2]$ and $Q_{h,i}$, for each sentence h in the answer of the i-th subject as:

$$Cr(W_{h,i,t1,t2}\ Q_{h,i}) = |\ W_{h,i,t1,t2} \text{ and } Q_{h,i}\ |\ /\ |W_{h,i,t1,t2}| \qquad (1)$$

Here, $W_{h,i,t1,t2}$ is the set of words expressing the component objects in the target image, which the i-th subject's eyes traced between time t_1 to t_2. For example, "*the actress*" and "*the perfume*" form $W_{h,i,t1,t2}$ if the i-th subject looks at the woman and the perfume bottle from the SSL from time t_1 to t_2.

If an object is included in $Q_{h,i}$, i.e., if an object has been referred to by the i-th subject, in his/her h-th unit of the answer, then the words in $Q_{h,i}$ are reused in expressing the elements of $W_{h,i,t1,t2}$. For example, if the subject called the woman in Fig. 8 "*the actress*," then this woman is put into the symbolic word "*the actress*" also in $W_{h,i,t1,t2}$. This is for enabling to take the conjunction and disjunction in Eq.(1), by calling the same object by the same name for both in $W_{h,i,t1,t2}$ and $Q_{h,i}$. If the object is not included in $Q_{h,i}$, words given before experiment were taken for the elements of $W_{h,i,t1,t2}$. For example, we may call the woman in Fig. 8 "Woman_1" and use this in $W_{h,i,t1,t2}$ for meaning the woman.

We evaluated the value of *SSL-likeliness* of $Q_{h,i}$ defined by $max_{\ t1,t2}\ Cr(W_{h,i,t1,t2}, Q_{h,i})$, i.e., the correspondence of $Q_{h,i}$ and the SSL corresponding most strongly to $Q_{h,i}$. Then, we counted the number of $Q_{h,i}$ for each subject s_i, of which *SSL-likeliness* is 1. In other words, we counted the words, sentences, and comments about target images, corresponding with one of the subject's SSLs.

For example, if the h-th sentence in the answer of subject s_i for the second questionnaire is "*I looked at the Japanese woman with the perfume*," and if there is an SSL between time t_1 and t_2 connecting the woman and a perfume bottle, as in Fig. 10, the value of *SSL-likeliness* of $Q_{h,i}$ defined by this sentence is 1. On the other hand, if the sentence is "*An actress with the man behind*," then we do not count this as an SSL-like sentence because "*at the man behind*" replacing "*the perfume*" reduces the *SSL-likeliness* of $Q_{h,i}$ down to 1/2, since the SSL in Fig.10 does not touch the men behind.

Results

Let us take one subject (call him Subject 1) and his responses to the questionnaires. In the first cycle (seeing image 1 through 5 and answering the questionnaire), the eye track on image 1 was as in the dots in Fig. 11. Subject 1 recollected all images, of which Poster 2 was recollected first, i.e., this was the most impressing to him. And, Poster 1 was in the second. In this questionnaire, his comment on image 1 was "*You are showing a poster of a perfume I once saw. The surrounding lights look like a night club. The woman is nice!*" for the questionnaire after seeing the first 5 posters.

The first sentence does not include any object in the image his eyes traced, so its SSL-likeliness is 0. The second and the third sentences include the underlined words corresponding to the component objects of the image. On the other hand, the three conductors of the experiment judged "night club" is not an object in the poster, but a metaphorical statement about the impression of the background scene. However, on the definition of correspondence give in Eq.(1), which means the fitting of an eye-track line to whatever part of a word sequence, both the *SSL-likeliness* of the second and the third sentences are 1. Therefore, these sentences are counted as SSL-like units.

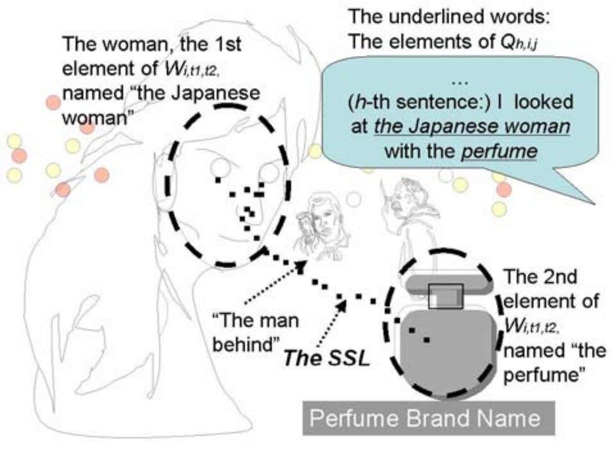

Figure 10. An SSL-like $Q_{h,i}$, i.e., the *SSL-likeliness* of $Q_{h,i}$ is 1.

In the second cycle (seeing Poster 6 through Poster 10), his comment on Poster 6 (same as Poster 1) changed to "*You show the perfume again! However, I newly recollect the woman is an actress who played in a famous movie.*" In this case, his eye track was as in the previously shown Fig. 8. The first sentence includes "perfume" corresponding to one SSL as in the small square on the bottle in Fig.8. And, "*woman*" in the second sentence corresponds to another SSL, so the *SSL-likeliness* of both sentences are 1. For Subject 1, in both cycles, the underlined *words* corresponded to SSLs.

On the other hand, another subject (Subject 2) had the eye track as in Fig. 12 for the first look at Poster 1. She said "*The poster of the perfume A with actress B*," where *A* and *B* were the real names of the perfume and the actress respectively. The long SSL in the square frame of Fig. 12 exactly corresponds to the combination of two words *A* and *B*. The *SSL-likeliness* of this sentence is 1, whereas the *SSL-likeliness* of each word *A* and *B* are 0.5. On the other hand, when she looked at the same picture for the second time (as Poster 6), her eyes traced as in Fig. 13, where we do not find an SSL connecting the woman's face and the perfume bottle. She said "*This is perfume B.*" The *SSL-likeliness* of this sentence and of the word "perfume B" are both 1.

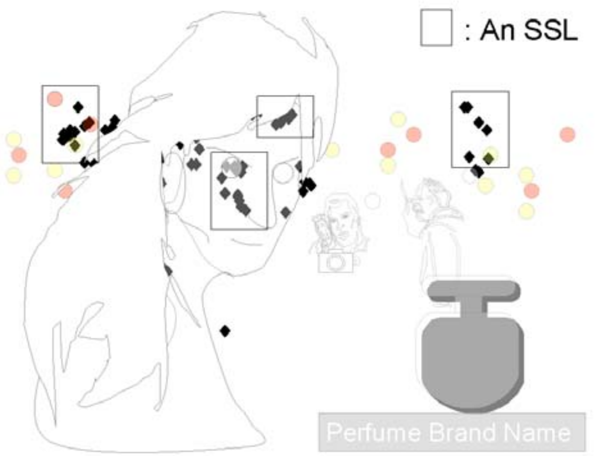

Figure 11. Subject 1's eye tracks in looking at Poster 6, i.e., the second appearance of Poster 1.

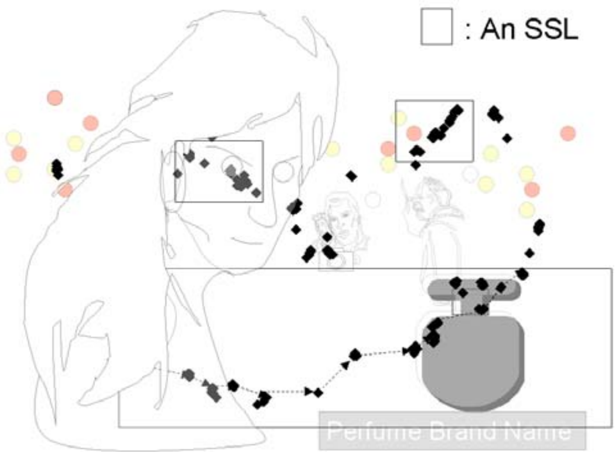

Figure 12. Subject 2's eye tracks for Poster 1

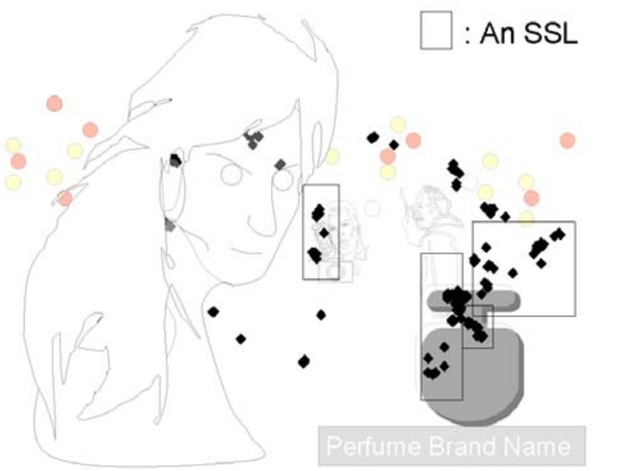

Figure 13. Subject 2's eye track for Poster 6, i.e., for the second appearance of Poster 1.

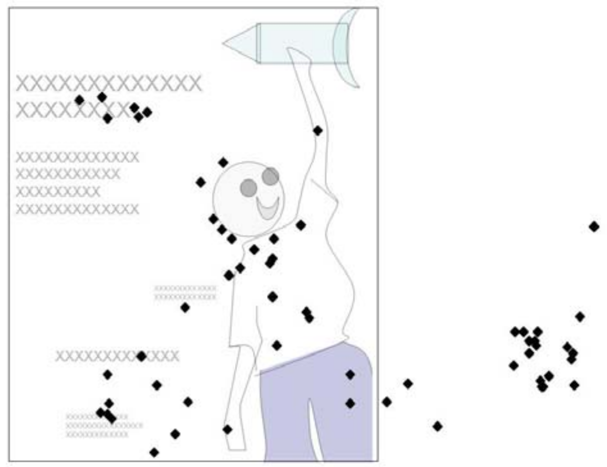

Figure 14. Subject 3's eye track for Poster 4

On of the exceptional cases where the SSL-likeliness of no sentence was equal to 1, was as in Fig. 14. The comment of this subject (Subject 3) was "*The boy has a toy rocket*" in which the "*rocket*" is obviously referring to the toy in the upper right of Poster 4, although the eye track do not reach the toy. In this case, we also found the eyes moved to outside of the poster, which may be due to the subjects' thinking about different things than this experiment.

The results are shown in Table II. According to the definition in Eq.(1) of correspondence degree between a word-sequence and an eye-track line, reader may expect the correspondence increases if we take a longer sequence as a unit $Q_{h,i}$ because $|\ W_{h,i,t1,t2}$ and $Q_{h,i}|$ increases with the size of $Q_{h,i}$ whereas $|W_{h,i,t1,t2}|$ does not change. In other words, a comment includes all sentences the subject wrote about the image, so a comments' SSL-likeliness is 1 as far as one of its sentences is of SSL-likeliness equal to 1. This stands also for a sentence and a word, i.e., a sentence's SSL-likeliness is 1 as far as one of its words is of SSL-likeliness equal to 1. According to Table II, however, the difference between the SSL-likeliness of sentences and of words is incomparably larger, than the difference between sentences and comments. From this result, we regard an SSL corresponds to a sentence, which we extracted as a semantic unit corresponding to an impression in each subject's comment.

Furthermore, in the answer to the questionnaire, the images of which the units of word-sequence did not match SSLs highly, i.e., words, sentences, or comments of low *SSL-likeliness*, were ranked lower than those of high *SSL-likeliness* comments, as in Table III. In this table, we computed average and standard deviations for comments including at least one meaningful word for the experiment (i.e., we ignored such comments as "this experiment is boring" "I am hungry" or empty blanks in the questionnaire answering form). This means that an image tends to be recollected well, if the content of an interpretation corresponds to SSLs.

TABLE II.
THE RATE OF WORDS, SENTENCES, AND COMMENTS, OF SSL-LIKELINESS EQUAL TO 1, AMONG ALL IN THE ANSWERS FOR THE 10 IMAGES, FOR EACH SUBJECT.

subject	The rate of SSL-like words	The rate of SSL-like Sentences	The rate of SSL-like Comments
s_1	0.16	0.63	0.66
s_2	0.13	0.78	0.78
s_3	0.67	0.89	0.90
	. .	. .	
s_{20}	0.20	0.93	0.96
s_{21}	0.29	0.75	0.76
Average (std)	0.33(.35)	0.80 (.29)	0.82 (.28)

TABLE III.
THE SSL-LIKELINESS OF WORDS, SENTENCES, AND COMMENTS, IN THE ANSWERS FOR THE 10 IMAGES, FOR EACH SUBJECT: SHOWING AVERAGE (STANDARD DEVIATION) IN EACH CELL.

subject	SSL-likeliness of words for image objects	of Sentences	of Comments
Rank 1	0.33 (0.23)	**0.92** (0.24)	**0.95** (0.22)
Rank 2	**0.35** (0.25)	0.91 (0.30)	0.91 (0.30)
Rank 3	0.29 (0.18)	0.83 (0.33)	0.86 (0.36)
Rank 4	0.23 (0.13)	0.75(0.33)	0.76 (0.34)
Rank 5	0.26 (0.15)	0.78 (0.35)	0.780.37)

B. Comparison with questionnaire one week later

Let us show the results of the questionnaire we took 1 week after the experiment, asking the same thing "write freely about the posters you can recollect, in the order you recollect easily" as in the questionnaire we conducted just after the subjects viewed the images. This time we did not show the images

In this questionnaire, among 21 subjects, between 1 and 8 subjects could recollect and speak about their interpretations or impressions for each image, whereas between 16 and 21 subjects did just after looking in Section A. However, the strength of their memory corresponded keenly with the appearance of SSLs, appearing just after the eyes moved in the velocity of saccade, i.e., between 60 and 600 deg/sec, as in Fig. 15.

If the memory of human decays with time monotonously, we can expect images shown later should be recollected well. Only in the case of some special impression, the earlier presented images might be recollected strongly. However, according to Fig. 15, the number of subjects who recollected each image decreased at the 4th, 5th, and the 8th images, among which the 4th and the 8th co-occurs with the decrease in the emergence of SSLs. And, considering the fact that the 1st and the 6th images were the same, i.e., the image shown twice, the increase in the number of recollecting subjects at the 2nd and the 7th image is counter intuitive. However, these match with the decrease in the emergence of SSLs. Among the 5 images of which the number of recollecting subjects increased from ones shown just before, the emergence of SSLs also increases for 4 images.

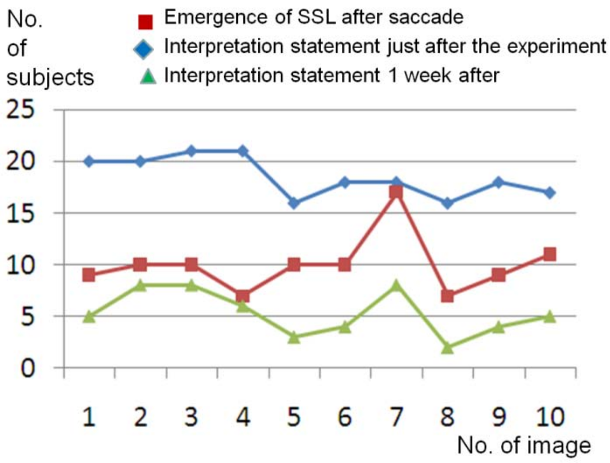

Figure 15. The tendency of long-term memory of interpretations, to correspond to the appearance of SSLs *just after saccade*.

C. Discussion

In subsection *A*, from the correspondence of SSLs and the sentences in the on-site questionnaire i.e., just after viewing each image, we showed an SSL corresponds to the essential unit of image interpretation. This is because each sentence we took from the questionnaire answer was a description unit of interpretation which remains in the subject's impression. Intuitively, the SSL draws an auxiliary line for connecting objects in the image to compose a meaningful story to be perceived as one's interpretation of the target image.

We could not say, however, that the hypothetical cognitive model in Fig. 2 has been validated on the evidences presented in subsection A. That is, the appearance of an SSL just after eyes explore hidden structures, as in Step 2' and Step 3' in Section II ((2) and (3) in Fig. 2), has not been observed in subsection *A*.

We found a clue to show the existence of this ordered steps in subsection *B*, i.e., from Step 2' to Step 3', from the questionnaire results of one week later. That is, the emergence of SSLs (auxiliary lines) just after the eyes move in the speed of saccade (search) corresponds to the impression strength of the interpretation in the long term memory of 1 week after viewing the image, according to the result of subsection *B*.

VI. CONCLUTIONS

To summarize the results in conclusion, we can revise the human's interpretation process of an image, as we showed in Fig. 2, into Fig. 16. That is, the slow saccade line (SSL) in (3) (Step 3' in Section II) appears as an auxiliary line, not only after fast saccade lines in Step (2) (Step 2' in Section II) but also occasionally for inspirations by chance. However, the emergence of SSLs just after fast saccade, following these two steps, tends to enable the interpretation on the auxiliary line to be acquired as long term memory. Therefore, we should not say "the viewer runs this process in image interpretation," but should conclude "if the viewer follows this process, he/she acquires a new interpretation of the image in his/her long term memory."

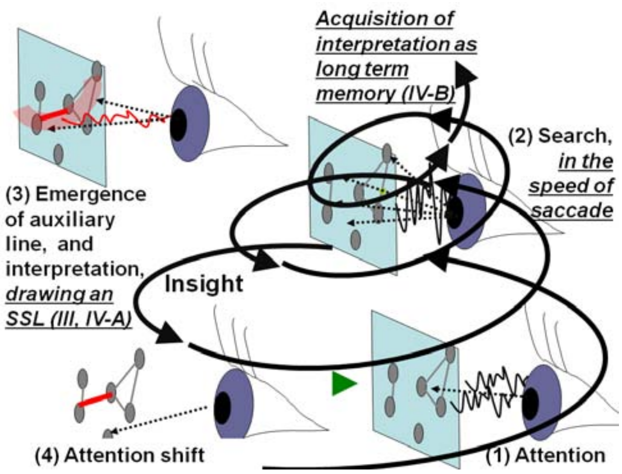

Figure 16. SSLs related to the process of image interpretation, and to the acquisition of interpretation as long term memory

REFERENCES

[1] Pieters, R., Warlop, L., "Visual Attention during Brand Choice: The Impact of Time Pressure and Task Motivation," *Int'l J. of Research in Marketing,* 16 (1), pp.1-16, 1999.

[2] Pieters, R., Warlop, L., and Wedel, M., "Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity for Brand Attention and Memory," *Management Science*, 48 (6), pp.765-781, 2002

[3] Pieters, R., and Wedel, M., "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing,* 68 (2), pp.36-50, 2004

[4] Just, M. A., and Carpenter, P. A., "A theory of reading: From eye fixations to comprehension," *Psychological Review*, 87, pp.329-35, 1980

[5] Rayner, K., "Eye movements and cognitive processes in reading, visual search, and scene perception," In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye Movement Research: Mechanisms, Processes, and Applications*. New York: Elsevier Science, 1995

[6] Salvucci, D. D. & Anderson, J. R., "Tracing eye movement protocols with cognitive process models," *Proc. of the Twentieth Annual Conference of the Cogniive Science Society*, pp. 923-928. Hillsdale, NJ, 1998

[7] Terai, H. and Miwa, K., "Insight problem solving from the viewpoint of constraint relaxation using eye movement analysis," *Proc. of the 4th international conference of cognitive science*, pp.671–676, 2003

[8] Knoblich G.; Ohlsson S.; Raney G. E., "An eye movement study of insight problem solving," *Memory & Cognition*, 29 (7), pp. 1000-1009, 2001

[9] Ohsawa, Y., and Nara, Y.: "Understanding Internet Users on Double Helical Model of Chance Discovery Process" *New Generation Computing*, Vol.21 No.2, pp.109-122, 2003

[10] Ohsawa, Y., "Chance Discovery: The Current States of Art," Ohsawa, Y., and Tsumoto, S. (eds), *Chance Discoveries in Real World Decision Making*, pp.69-82, Springer Verlag, 2006

[11] Horie, K., and Ohsawa,Y., "Product Designed on Scenario Maps Using Pictorial KeyGraph," *WSEAS Transaction on Information Science and Application*, Vol.3 No.7, pp.1324-1331 (2006)

[12] Suwa, M., and Motoda, H., "Frustration-Based Learning in Auxiliary-Line Problems in Elementary Geometry," J*ournal of Japanese Society on Artificial Intelligence* 4(3), pp.308-320, 1989.

[13] Matsuda, N.., and Okamoto, T.., "Diagrammatic Reasoning for Geometry ITS to Teach Auxiliary Line Construction Problem," In B. P. Goettl, H. M. Halff, C. L. Redfield & V. J. Shute (Eds.), *Proc. of the International Conference on Intelligent Tutoring Systems*, pp.244-253, Heidelberg, Berlin: Springer, 1998

[14] Norton, D., and Stark, L., "Eye Movements and Visual Perception," *Scientific American* 224, pp.34-43, 1971

[15] Martinez-Conde, S.,. et al., "The role of fixational eye movements in visual perception," *Nature Reviews Neuroscience* 5, 229-240, 2004

Yukio Ohsawa is an associate professor in the School of Engineering, The University of Tokyo. He received Ph.D in Communication and Information Engineering from The University of Tokyo. He worked also for School of Engineering Science in Osaka University (research associate, 1995-1999), Graduate School of Business Sciences in University of Tsukuba (associate professor, 1999-2005), and Japan Science and Technology Corporation (JST researcher, 2000-2003). He initiated the research area of Chance Discovery, defined "discovery of events significant for decision making" in 1999, and series of international meetings (conference sessions and workshops), e.g., the fall symposium of the American Association of Artificial Intelligence (2001). He edited the first book on "Chance Discovery" (2003) and "Chance Discoveries in Real World Decision Making" (2003) published by Springer Verlag, and special issues in international and Japanese (domestic) journals. Chance discovery is growing: Journal issues has been published from the international journals. He is in the editorial board the Japanese Society of AI and the planning board of New Generation Computing, and is the TC chair of IEEE-SMC technical committee of Information Systems for Design & Marketing.

Yusuke Maeda graduated from Tokyo University of Science (2006), and is a graduate student in the School of Engineering, The University of Tokyo.