ESE 356/501 System Specification and Modeling
Department of Electrical and Computer Engineering
Stony Brook University

Gaze Behavior Monitoring System

Fig. 1 illustrates the basic concept of the proposed application model. When a visual content is provided, the provider naturally reflects the intention to the content in order to accomplish the goal of the content. For example, if it is an education content, the provider might try to emphasize key content and core sequence. If it is a commercial content, the provider might want a consumer to concentrate one's attention on not an advertising model but a main product. Thus, the provider deliberately generates the content in the several aspects such as showing object, sequence, and duration time. It is the intended pattern B as shown in the figure.

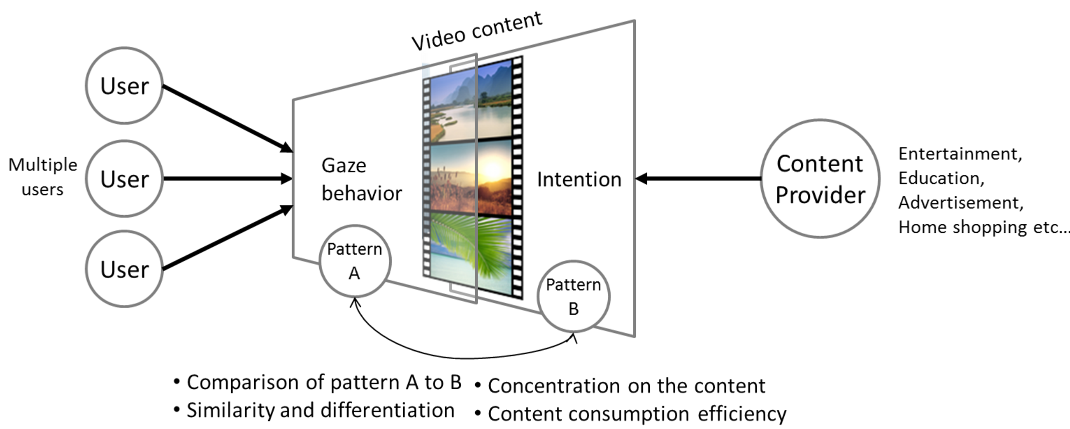

Fig. 1 Conceptive illustration of the proposed application model

On the other side, when multiple users gaze at the multimedia content, there are numerous patterns of gaze behavior. It is the pattern A of gaze behavior as shown in the figure. Pattern A is different depends on the user. In addition, even though it belongs to the same user, the pattern is changed by the perceived context like usage environment including a time slot and a place. The important thing is that the gaze behavior is affected by the intention of the content based on visual object and audio effect. If the intentional visual and audio information have less of an effect on the gaze behavior, it implies that the user doesn't concentrate on the content. Thus, by doing comparison of pattern A to B, the degree of content usage efficiency such as concentration and discursiveness can be measured as a useful information to improve the service quality.

3.1 Intention Representation and Annotation

Intention of the content is created by the provider, which is simply a flawless footprint for user to follow of the content usage. It shows that what is more important content and what is the best order to use the content considering time flow. In order to reveal these intentions, the visual content has to be able to represent the intentional condition including intentional sequence and weight. Therefore, in this paper, intention is defined with two criteria: intention flow and intention weight.

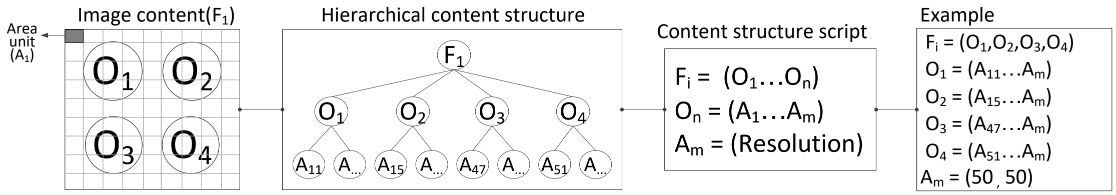

Fig. 2 Illustration of the content hierarchical structure and the representation script

An individual object information should be embedded within the image as a precondition to express the intention. Fig. 3 illustrates the hierarchy structure of the image content from unit area (A_m) to object (O_n), and to image frame (F_i). An object has at least more than one unit area. By using this structure, the video image content structure script made up of F_i, O_n, and A_m is generated. When F_1 has O_1, O_2, O_3, and O_4, the script of F_1 describes it like the simple example in the figure.

Fig. 4 shows the abstractive gaze information on the basis of an image content. When a user looks at the image, the gaze data is captured as a single point unit. In addition, an individual object is represented by the area information, which is made up of the group of points. Thus, the point information is a common element of the gaze and object. They are combined with each other according to the relationship between the point and the area of the object. If a pixel is chosen during the gaze tracking, the corresponding object can be obtained. In order to annotate the gaze data with the object, the object based content structure representation model described in Fig. 3 is applied in this step.

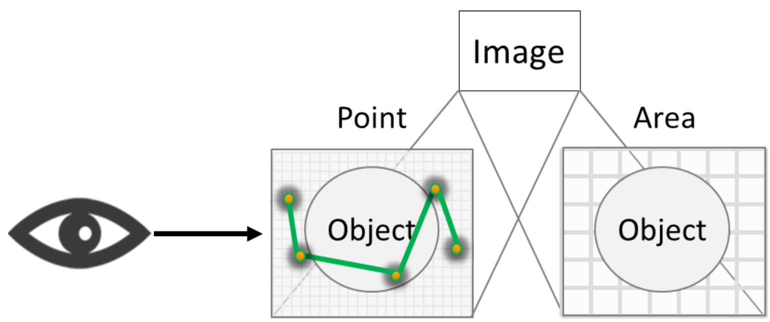

Fig. 3 Abstractive gaze information on the basis of an image content

Fig. 5 explains the gaze capturing and temporal annotation approach in order to compute the gaze duration. When the system captures gaze data, it combines gaze point with the object information using the content structure script. The gaze duration is calculated by the number of captured gaze fixations within the object, and sampling time. After initial capturing, the gaze profile is generated as a positional data based on the point, temporal data based on the sampling time, and associated object data based on the content structure script. Through the profiling process, the system keeps the gaze duration and object based gaze sequence.

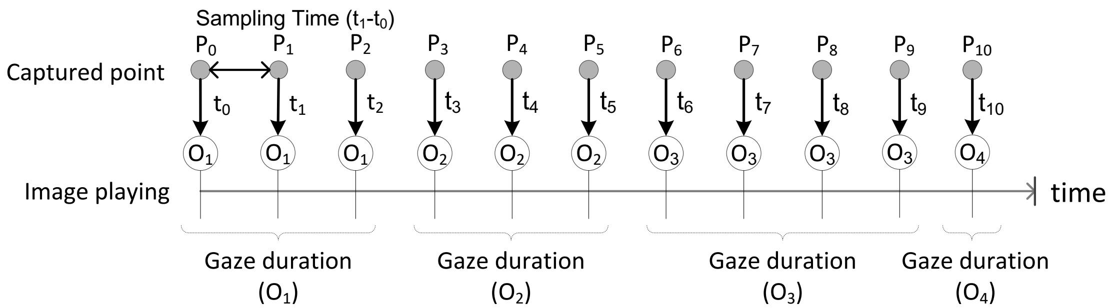

Fig. 4 Gaze duration using data capturing and annotation

After the gaze profile has obtained, it is possible to draw a comparison between object based intention flow and gaze profile. Fig. 6 shows the example of both of the data for comparison. In the figure, the intentional object ordering is O_1, O_4, O_2, and O_3. But gazing behavior has annotated O_1, O_2, O_3, and O_4 in order. If the intention is not considered, there is no way to judge the gaze behavior is good or bad. However, this gaze behavior is not perfect in terms of the intention because of the difference order within the sync segment. It is just simple case, thus, the data stored through the gazing data annotation would take more important role when the diverse intention weight is applied for comparison.

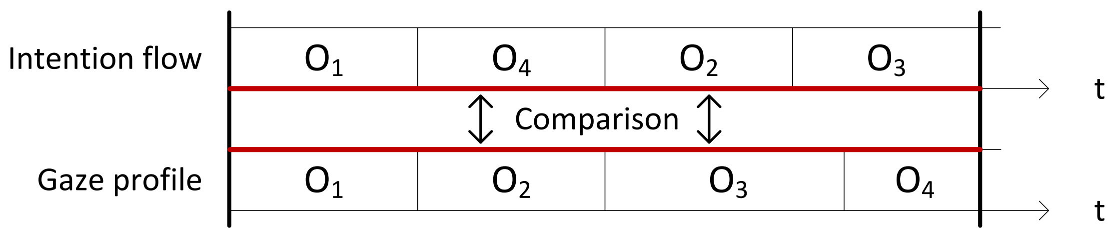

Fig. 5 Illustration of the comparison intention flow with gaze profile

The temporal information determines the intentional sequence of the object. In the image content, the temporal information is synchronized by an audio information. When a content provider combines the visual data with audio such as narration or music, there is a sync segment which indicates the specific time that the object and audio data is match up to signify the intention. Fig. 7 describes the intentional sequence based intention flow script. When an audio data is available to the content, the provider creates the audio sync segment, AS, to associate the object with audio data based on time. Content provider can flexibly define the range of the intention using AS.

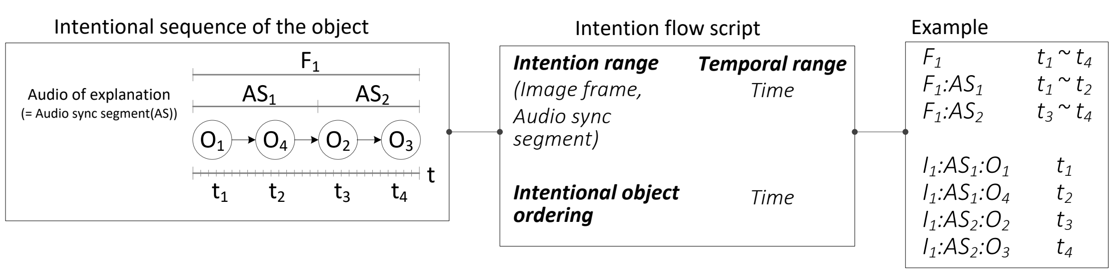

Fig. 6 Illustration of the intentional sequence of the object and intention flow script

Thus far, the intention is only represented by sequences of objects ordering. However, all of the objects have the same importance due to the purpose of the content. In order to define the relative object importance, the intention weight script is proposed. Fig. 8 shows the structure of intention weight script based on intentional importance of the object, group, and sequence condition. For example, a user who looks at the object which is not important, O_1, and a user who looks at the object which is very important, O_4, O_2, O_3, are quiet distinct from each other. In order to disclose this distinction, the weight condition represents intentional target with relative importance weight value. When a new condition is defined, it also has to be indicated the weight. When the provider sets up the weight values, the weight value assignment strategy is required according to characteristics of the target because of the possibility of reverse result in terms of values. In other words, a user who looks at the targeted group has to have higher than a user who looks at a little object among the group.

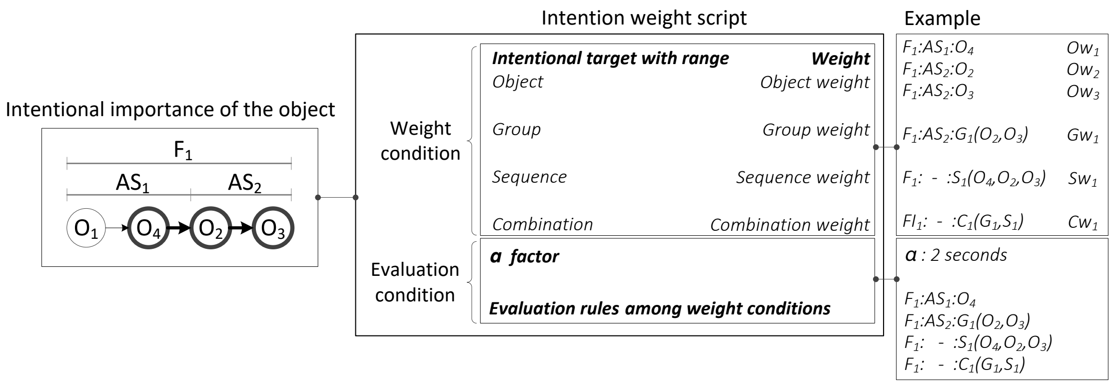

Fig. 7 Illustration of the intentional importance of the object and intention weight script

In order to handle this, weight of group (Gw) has higher weight than individual weight of object (Ow). It means, $G_{(O1, O2)}$ has higher importance than individual object O_1, and O_2. Also, it is possible that O_1, O_2 in order is important than O_2, O_1 because of the intent of the provider. So, in this case, sequence event satisfaction is more important than a group. In order to assign the sequence weight value, weight of sequence (Sw) has higher weight than Gw. Moreover, complicated conditional intent is possible. In order to use the combination weight, weight of combination (Cw) has higher weight than individual condition, for example, $Cw(G_1 \mid S_1) > S_1$. The reason why complex condition has more weight than a single condition is because that the relative importance of data can be increased when the complex condition is satisfied.