動画共有システムにおける ユーザコメントの領域と時区間を用いたシーン抽出

若宮　翔子† 　北山　大輔†† 　角谷　和俊†

† 兵庫県立大学環境人間学部　〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12
†† 兵庫県立大学大学院環境人間学研究科 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12
E-mail: †{nc05q204,ne07p001}@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

あらまし　近年，インターネット上に動画を投稿し，他のユーザと共有することができる動画共有サイトが注目されている．このようなサイトにおいて，ユーザは動画に対するコメントを投稿することができる．しかし動画に対するコメントは大量である上に，それらの関係が不明であり，またユーザによって視聴したいシーンが異なるため，ユーザにとって不要なコメントやシーンが存在する．そこで我々はユーザの興味に関連するシーンを，ユーザが選択したコメントから抽出する手法を提案する．本研究では，ユーザが画面領域と時区間を指定して投稿したコメントを利用できる環境を想定している．そしてコメントとその画面領域によって表されるオブジェクトと，コメントとその時区間によって表されるイベントに着目し，画面領域の位置関係により同一オブジェクトを，時区間の前後関係により同一イベントを判定し，それらの判定結果を組み合わせてシーン間の関係を定義し，関係に基づきシーン抽出を行う手法を提案する．本稿では，オブジェクトとイベントの判定方式とそれらの判定結果を組み合わせたシーン間の関係，提案手法に基づき開発したプロトタイプシステムについて述べる．

キーワード　マルチメディア，動画共有，コメント，画面領域，時区間，シーン抽出

Scene Extraction for Video Sharing based on the Relation of Text, Pointing Region and Temporal Duration of User Comments

Shoko WAKAMIYA†, Daisuke KITAYAMA††, and Kazutoshi SUMIYA†

† School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan
†† Graduate School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan
E-mail: †{nc05q204,ne07p001}@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

Abstract　Recently, video sharing websites that allow users to attach comments to videos have attracted much attention. In this paper, we propose a method whereby users can easily retrieve video scenes relevant to their interest. Our system makes both a text and non-text analysis of a user's comment and then retrieves and displays relevant scenes for viewing of the scenes along with attached comments. The text analysis works in tandem with non-text features, namely, the pointing region and temporal duration of user comments. In this way, our system supports a better organized retrieval of scenes that have attached user comments with a higher degree of relevancy for the user than is currently available with conventional methods, for example, using matching keywords. We describe here our method and relations between scenes and discuss a prototype system.

Key words　Multimedia, Video sharing, Comment, Pointing region, Temporal duration, Scene extraction

1. Introduction

Recently, the number of video sharing websites has rapidly increased, and a lot of videos are being shared online. Users can write comments about a shared video with applications like BBS interface on various sites such as YouTube [1] and

Google Video [2]. On these sites, users often write comments about an entire video. However, users are able to write comments about a particular scene on sites such as Nico Nico Douga [3] and jimaku.in [4], and these comments that are spuriously synchronized with the playing of the videos are displayed on the screen for a definite time interval. This way users can easily write comments about shared videos at video sharing websites resulting in comments that are irrelevant to the user and often users viewing the same shared video want to view different scenes resulting in scenes that are also irrelevant to each user. For example, when a user views a video of a baseball game and is attracted to Ichiro in the video, scenes unrelated to Ichiro are not required by that user. And when the user views Ichiro's scene, comments unrelated to Ichiro are also not required by that user. Therefore, the extracting method of comments and scenes related to the user's interests is necessary. However, using the conventional method of matching keywords is not adequate to extract them because the relation among such comments and scenes are not clear with this method.

In this paper, we propose the extracting method of comments and scenes related to a user's interests based on the relation among other user comments. So, we are looking at a situation where users can write comments with a specified pointing region and temporal duration using the comment posting system we have developed (see Fig. 1). A user comment therefore consists not only of keywords, but also of a specified pointing region and temporal duration. Fig. 2 shows these user comments, and x axis and y axis are screen coordinates and t is the time axes. The text is treated as a keyword set consisting of a noun, verb, adjective and adverb. The pointing region is the specified area on a video screen selected by a user where the user wants to attach a comment to. It consists of coordinates and area size. The temporal duration is the specified time interval by the user which continues to display a comment and the pointing region of it. It consists of a starting point and an ending point. We divided a video into scenes based on image processing and speech recognition techniques of the current scene division systems [7] [8].

In our proposed method, users can view scenes and their attached comments concerning the object and event they are interested in, merely by selecting a comment of interest. Take, for example, Ichiro is an object and a home run is an event in a scene from a baseball game. The object and event are defined based on the comment's pointing region and temporal duration, as well as the comment's text. We determine the object by the relation of the selected comment's text and the pointing region and other comments. Then, determine the event by the relation of the selected comment's text and

図 1　コメント投稿システム
Fig. 1　Screen image of comment posting system

the temporal duration and other comments. We describe the advantage of our method as follows:

（1） **The advantage of the object determination method**　By using the pointing region, our method can locate scenes relevant to the user's comment even when the comment does not contain keywords found in other comments of the same object. Also by merit of using the pointing region, our system can determine where keywords common to comments in fact refer to different scenes.

（2） **The advantage of the event determination method**　Because videos are rarely annotated by the uploader, background information concerning particular events appearing in the video is not available. Even so our method can locate these events as comments are linked with temporal duration.

The remaining sections of this paper are as follows: Section 2 describes the object determination based on the pointing region. Section 3 describes the event determination based on the temporal duration. Section 4 describes the scene retrieval application. Section 5 discusses the evaluation, and Section 6 reviews related work.

2. Object determination based on pointing regions

When a user attaches a comment on the screen, the user watches an object and specifies its pointing region. An object is defined by keywords in the comment and the pointing region (see Fig. 3). By comparing keywords and the degree of the overlap of the pointing regions or the relative position of the pointing regions, we estimate whether an object indicated by the selected comment matches other comments.

$$C_{object} = \{c_o | Cov(c_s, c_o) \geqq \alpha \ or \ Pos(c_s, c_o) \geqq \beta\} \quad (1)$$

c_s is a user selected comment and c_o is an extracted com-

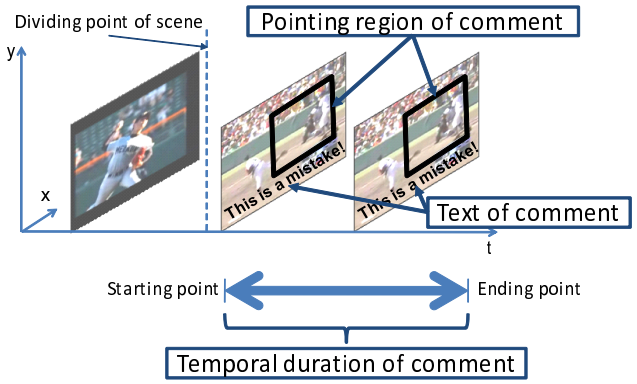

図 2　指定画面領域と指定時区間を持つコメント
Fig. 2　Comment with user pointing region and temporal duration

ment indicating the same object as c_s. The function Cov calculates the degree of the overlap of the pointing regions, and the function Pos calculates the relation of the relative position of the pointing regions. If Cov or Pos is higher than the threshold, we detect c_o as the related comment about the same target object.

This object determination method has two advantages. First, our method can estimate a comment as indicating the same object as the selected comment even when the comment does not contain keywords found in other comments of the same object. For example, when a user watches a shared video of a soccer game and selects the comment "Nice save" that specifies the pointing region of a goalkeeper on the video screen, our method using the pointing region can extract the comment "The keeper's reaction is good" as indicating the same object as the selected comment even if it does not contain the keywords "nice" and "save". Second, our method can extract a different object from common keywords found in the selected comment. For example, there are two goalkeepers in a soccer game, and they are often expressed with the same keyword, "keeper". When a user selects the comment "The keeper's adrenalin is pumped!" that specified a pointing region of team A's goalkeeper on the video screen, our method using the pointing region can filter out the comment "This keeper is tough" as indicating the different object, namely, team B's goalkeeper as the selected comment even if it contains the keyword "keeper".

2.1 Object determination method using degree of overlap of pointing regions

The degree of the overlap of the pointing regions is high in videos with still or steady images, and so these tend to be specified by close to absolute positions. The degree of the overlap of the comments' pointing regions will tell us if the objects they refer to are the same or not. We calculate the function Cov as follows:

$$Cov(c_s, c_i) = CovR(c_s, c_i) \times (CovK(c_s, c_i) + \gamma) \quad (2)$$

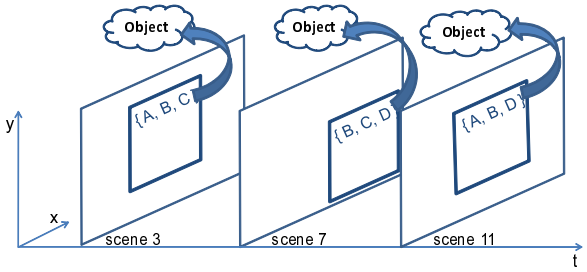

図 3　オブジェクトの概念図
Fig. 3　Concept image of object

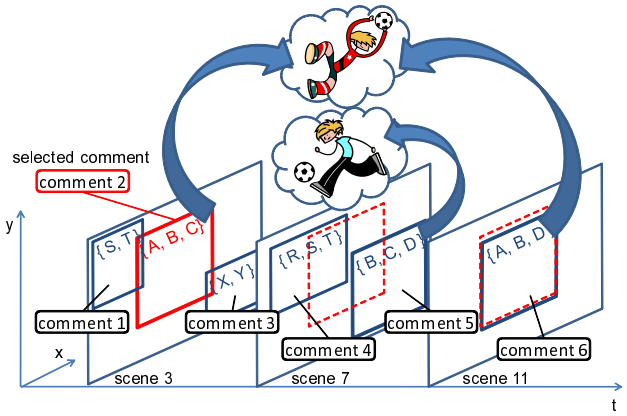

図 4　領域の重複度を用いたオブジェクト判定の例
Fig. 4　An example of object determination using degree of overlap of pointing regions

$$CovR(c_s, c_i) = \frac{|R(c_s) \cap R(c_i)|}{|R(c_s) \cup R(c_i)|} \quad (3)$$

$$CovK(c_s, c_i) = \frac{|K(c_s) \cap K(c_i)|}{|min(K(c_s), K(c_i))|} \quad (4)$$

when c_s is a user selected comment and c_i is one of other comments. The function $CovR$ calculates the degree of the overlap of the pointing regions. The function $CovK$ calculates the similarity of keywords. γ is a constant that prevents $CovK$ from becoming 0.

When a user selects the comment "Nice save" that specifies the pointing region of a goalkeeper on a video screen, our method using the pointing region can extract comments as indicating the same object as the selected comment even if they do not contain the keywords "nice" and "save". Fig. 4 shows an example of the object determination using the degree of the overlap of the pointing regions. When the user selects comment 2 in scene 3, we extract comment 6 in scene 11 as indicating the same object as the selected comment. Because both the degree of the overlap of the pointing regions and the similarity of keywords between the selected comment and the other comment are high.

2.2 Determination method using relative position of pointing regions

By using the relative position of the pointing regions of

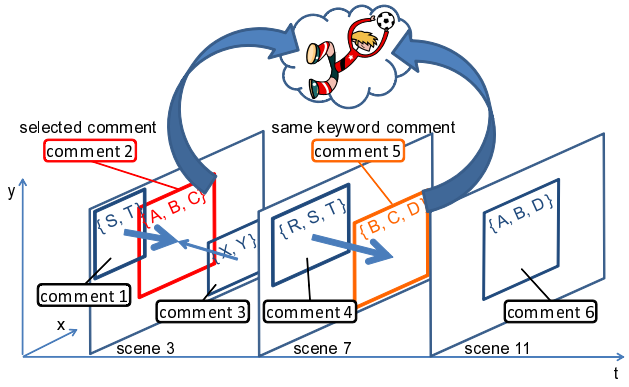

図 5　領域の位置関係を用いたオブジェクト判定の例

Fig. 5　An example of object determination using relative position of pointing regions

comments in the same scene, we determine whether or not the comments refer to the same object. This is effective when objects on the screen are constantly moving. Fig. 5 shows an example of the object determination using the relative position of the pointing regions. When the user selects comment 2 in scene 3, we extract the relative positions of the pointing regions of the selected comment and related comments, comment 1 and comment 3, from scene 3. And we also extract the same keyword comment, comment 5, and another related comment, comment 4, from scene 7. In this case, the selected comment is located to the lower right of comment 1 and to the upper left of comment 3. And comment 5 is located to the lower right of comment 4 in scene 7. Here, the lower right proximity of the pointing regions is determined for comment 1 and comment 4. So, we extract comment 5 as referring to the same object as the selected comment.

$$Pos(c_s, c_c, c_i, c_j) = RelP(c_s, c_c, c_i, c_j) \times covK(c_s, c_i) \times covK(c_c, c_j) \quad (5)$$

$$RelP(c_s, c_c, c_i, c_j) = \begin{cases} 1 \ (P(c_s, c_i) = P(c_c, c_j)) \\ 0 \ (P(c_s, c_i) \neq P(c_c, c_j)) \end{cases} \quad (6)$$

c_c and c_j are comments which are linked with c_s and c_i in another scene. The function $RelP$ returns 1 when the relative position of the pointing regions is same.

3. Event determination based on temporal duration

We define an event as indicated by common keywords in comments attached to adjacent scenes (see Fig. 6). We estimate whether events in separate scenes are the same or not by calculating as follows:

$$C_{event} = \{c_e | RelT(c_s, c_e, c_{s_p}, c_{e_p}) \times PairK(c_s, c_e, c_{s_p}, c_{e_p}) = 1\} \quad (7)$$

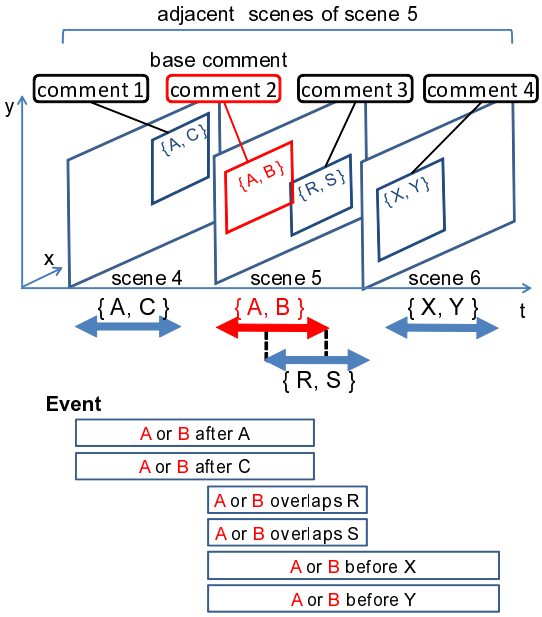

図 6　イベントの概念図

Fig. 6　Concept image of Event

c_s is a user selected comment and c_e is an extracted comment as indicating the same event as c_s. c_{s_p} is a comment which was attached to the scene of c_s, and c_{e_p} is a comment which was attached to the scene of c_e. If the function $RelT$ and $PairK$ returns 1, we detect c_e as a related comment referring to the same event.

$$RelT(c_s, c_{s_p}, c_i, c_{i_p}) = \begin{cases} 1 \ (T(c_s, c_{s_p}) = T(c_i, c_{i_p})) \\ 0 \ (T(c_s, c_{s_p}) \neq T(c_i, c_{i_p})) \end{cases} \quad (8)$$

$$PairK(c_s, c_i, c_{s_p}, c_{i_p}) = \begin{cases} 1 \ (c_s = c_i \ and \ c_{s_p} = c_{i_p}) \\ 0 \ (c_s \neq c_i \ or \ c_{s_p} \neq c_{i_p}) \end{cases} \quad (9)$$

$RelT$ returns 1 when the temporal durations' relationship of two keyword pairs is same. $PairK$ returns 1 when the two keyword pairs are same. We defined the temporal durations' relation using the relation types based on the relative position of temporal duration and Allen's time interval model [17] (see Table 1). In Table 1, t_{s_s} is the starting point of a selected comment and t_{s_e} is its ending point. t_{i_s} is the starting point of another comment and t_{i_e} is its ending point.

In Fig. 7, when a user selects comment 1 in scene 5, we extract the relation of the temporal duration between the selected comment and related comments, comment 2 and comment 3, in scene 5's adjacent scenes, comment 4 in scene 13 because this comment contains the same keyword "A", and related comments, comment 5 and comment 6, in scene 13's adjacent scenes. Then the extracted keywords in the comments are paired with the keywords in the selected comment making 8 different pairs. Also the extracted keywords in the comments are paired with the keywords in the comment containing the same keyword as in the selected com-

表 1　時区間の関係
Table 1　The relation of temporal duration

Relation type	Determination condition	Order
before	$\|t_{s_e} - t_{i_s}\| > 0$	$t_{s_s} < t_{i_s}$ and $t_{s_e} < t_{i_e}$
after	$\|t_{i_e} - t_{s_s}\| > 0$	$t_{s_s} > t_{i_s}$ and $t_{s_e} > t_{i_e}$
equal	$\|t_{s_s} - t_{i_s}\| = 0$ and $\|t_{s_e} - t_{i_e}\| = 0$	$t_{s_s} = t_{i_s}$ and $t_{s_e} = t_{i_e}$
meets	$\|t_{s_e} - t_{i_s}\| = 0$	$t_{s_s} < t_{i_s}$ and $t_{s_e} < t_{i_e}$
met-by	$\|t_{i_e} - t_{s_s}\| = 0$	$t_{s_s} > t_{i_s}$ and $t_{s_e} > t_{i_e}$
overlaps	$\|t_{s_e} - t_{i_s}\| > 0$ and $t_{s_e} > t_{i_s}$	$t_{s_s} < t_{i_s}$ and $t_{s_e} < t_{i_e}$
overlapped-by	$\|t_{i_e} - t_{s_s}\| > 0$ and $t_{i_e} > t_{s_s}$	$t_{s_s} > t_{i_s}$ and $t_{s_e} > t_{i_e}$
during	$\|t_{s_s} - t_{i_s}\| > 0$ and $\|t_{s_e} - t_{i_e}\| > 0$	$t_{s_s} > t_{i_s}$ and $t_{s_e} < t_{i_e}$
contains	$\|t_{s_s} - t_{i_s}\| > 0$ and $\|t_{s_e} - t_{i_e}\| > 0$	$t_{s_s} < t_{i_s}$ and $t_{s_e} > t_{i_e}$
starts	$\|t_{s_s} - t_{i_s}\| = 0$ and $\|t_{x_e} - t_{y_e}\| > 0$	$t_{s_s} = t_{i_s}$
finishes	$\|t_{s_e} - t_{i_e}\| = 0$ and $\|t_{s_s} - t_{i_s}\| > 0$	$t_{s_e} = t_{i_e}$

ment also making 8 different pairs. If a pair in the selected comment's pair and a pair in the keyword comment's pair are the same, this comment pair is determined as referring to the same event as the selected comment's pair. In this case, because the temporal duration's relationship of the pair "A" and "S" in the selected comment's pair and in the same keyword comment's pair is same, "overlaps", we estimate that these comment pairs are indicating the same event. Thus, the system searches for comments with the same keyword and temporal duration, and determines whether or not these comments refer to the same event.

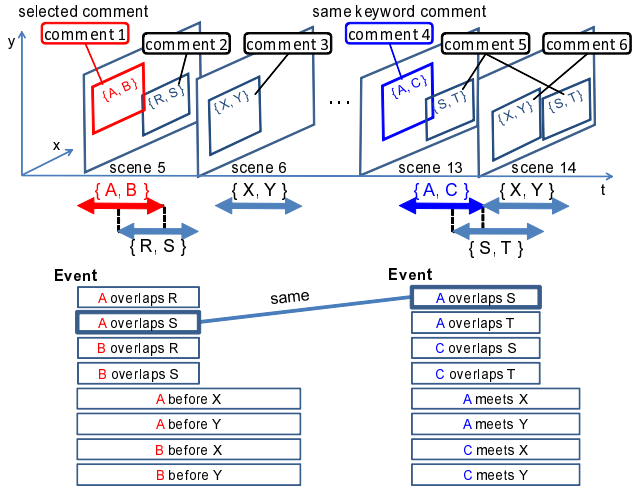

図 7　時区間の関係を用いたイベント判定の例
Fig. 7　An example of event determination using the relation of temporal duration

This event determination method has an advantage. Our method can extract comments as indicating the same event as the keyword pair in the selected comment and the other comment without background information concerning particular events appearing in the video. For example, when a user watches a shared video of a soccer game and selects the comment "Shoot!", our method using the temporal duration can extract the comment pair "Strong shoot" and "Nice keep" as indicating the same event as the selected comment and the other comment, "Keeper, nice reaction!". In this case, both comment pairs contain the keyword pair "shoot" and "nice", and we consider these comment pairs indicate the same event.

4. Scene retrieval application

We categorize scenes by type according to comments' objects and events. Scenes are then compared using this classification system and related comments retrieved accordingly.

4.1 Types of scene relation

We define 4 types of scene relation according to a comment's object and event as follows: *EQUAL*, *OBJECT*, *EVENT* and *NO RELATION*. Table 2 shows scene relation types. Other comments utilized in determining the event are included in the calculation, and the system determines whether or not the comments refer to the same object. We express the determination result of these comments 'Other-object' against the determination result of the selected comment and another related comment such as the same keyword comment 'Base-object' in Fig. 8 and Table 2. By using them, the system determines the relationship between the scene of the comment that a user is interested in and of the other.

EQUAL　The scene's relationship is defined as *EQUAL* when comments share the object and event.

OBJECT　The scene's relationship is defined as *OBJECT* when comments share both the base-object and event and other-object indicates the different object, or when the comments share the base-object only and other-object indicates the same object or different object.

EVENT　The scene's relationship is defined as *EVENT* when comments share the event only and other-object indicates the same object or different object, or when neither the base-object nor event is shared and other-object indicates the same object.

NO RELATION　The scene's relationship is defined as *NO RELATION* when neither the object nor event is shared.

表 2　シーン間の関係タイプ
Table 2　Relation types between scenes

		Object			
	Base-object	same	same	different	different
	Other-object	same	different	same	different
Event	same	*EQUAL*	*OBJECT*	*EVENT*	*EVENT*
	different	*OBJECT*	*OBJECT*	*EVENT*	*NO RELATION*

4.2 Scene retrieval concerning object

For a user who watches a shared video and is attracted to an object in the video scene, our proposed system is able

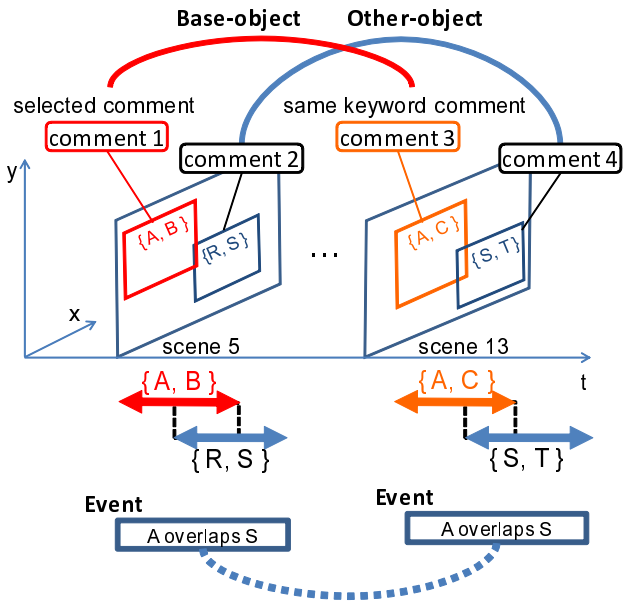

図 8　シーン間の関係判定の概念図

Fig. 8　Concept image of the determination of relation between scenes

to retrieve and present scenes of comments which refer to the same object that the user is interested in. When the user wants to retrieve scenes about the object of the user's selected comment, the system extracts thus scenes by using the scene relation $EQUAL$ and $OBJECT$. That is because we consider the user wants to view the same object in similar scenes as the selected comment's scene. For scenes with the relation $OBJECT$, the system extracts scenes with comments referring to the same object. For scenes with the relation $EQUAL$, scenes with comments determined to refer to the same object and same event in other scenes are extracted.

In our method by utilizing the relation of the pointing region and temporal duration, even if a comment does not contain the user comment's keywords, it will still retrieve relevant scenes if the comment refers to the same object. Our method can also determine when comments containing some of the same keywords refer to different objects. This is made possible by comparing the comments' pointing regions.

4. 3　Scene retrieval concerning event

As most shared video is not annotated, we consider it is useful for the user to be able to comment on and view scenes depicting an event of interest and thereby obtain more information of the event. Our system retrieves scenes including the event by comparing the comments' temporal duration.

For a user who watches a shared video and is attracted to an event in the video scenes, our proposed system is able to retrieve and present scenes of comments which refer to the same sequential event that the user is interested in. When the user wants to retrieve scenes about the event of the selected comment, the system extracts thus sequential scenes by using the scene relation $EQUAL$ and $EVENT$. That is because we consider the user wants to view the same sequential event in similar scenes as the selected comment's scene. For scenes with the relation $EQUAL$, scenes with comments determined to refer to the same object and same event in other scenes are extracted. Comments regarding different objects in the same event are located through a comparison of the $EVENT$ type relation. The system retrieves scenes depicting the same event even when comments about the event refer to different objects.

5.　Evaluation

5. 1　Prototype system

We developed a prototype system based on our proposed method using Microsoft Visual Studio 2008 C#. This prototype system consists of a video-viewing interface, a comment posting interface and lists of relevant comments concerning object and event. Fig. 9 shows a screen image of the prototype system. In this system, the user's selected comment is decided by selecting a comment which is displayed on the video-viewing interface, or by writing a comment to a specified pointing region and temporal duration using the comment posting interface. Then, the system extracts the comments related to the selected comment based not only on the text or keywords that are divided by the morphological analyzer Mecab [5] that is in SlothLib [6], but also on the pointing region and temporal duration, and displays the relevant scenes that the extracted comments are attached to.

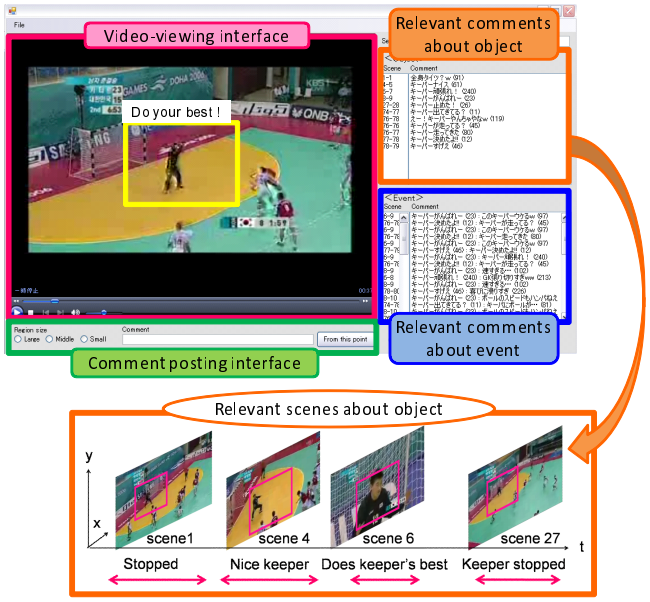

図 9　プロトタイプシステム

Fig. 9　Screen image of prototype system

5.2 Experimental evaluation

We used videos with comments that specified the pointing region and temporal duration as video data of the experiment. The videos we used were being shared in YouTube [1] and Nico Nico Douga [3]. There were about two hundred new comments per video that were attached using the comment posting system we developed (see Fig. 2). In this experiment, we gave each of the 4 subjects a list of some comments, and then they chose the comments and scenes they were interested in. Table 3 shows the precision, recall and F-measure about the scenes that were extracted by the prototype system based on our method. Also we evaluated the scenes that were extracted by using matching keywords in order to compare with our method (see Table 4). We got similar results about scenes related to object although the results of the comparison look similar, they are misleading because our method extracts comments not only based on keywords.

表 3　プロトタイプシステムにより抽出されたシーンの評価
Table 3　Evaluation of the extracted scenes by prototype system

	Scenes related to object	Scenes related to event
Precision	0.30	0.43
Recall	0.60	0.26
F-measure	0.40	0.32

表 4　コメントのキーワードの一致により抽出されたシーンの評価
Table 4　Evaluation of the extracted scenes by matching keywords

	Scenes related to object	Scenes related to event
Precision	0.28	0.52
Recall	0.58	0.35
F-measure	0.38	0.42

5.3 Examples of extracted scenes

5.3.1 Example of scenes related to object

We explain an example of scenes that are extracted as the result of the scene retrieval concerning object. In Fig. 10, when the user watches the shared video of the handball game and selects the comment “Doing various movements in a moment” that was written to the goalkeeper about the shot event in scene 12, our system extracts 2 other scenes as relevant scenes. Scene 8 and scene 66 are extracted because the scene relation is $OBJECT$, or there are comments that determined as the same object by the degree of the overlap of the pointing regions, “Strange movement” and “Wide movement” in scene 8 and “Strange pose” in scene 66. Also scene 8 is extracted because the scene relation is $EQUAL$, or there is the comment pair “Wide movement” and “GK is so excited” that determined the same event as the selected comment and the other comment “Come on, GK!” in scene 12. This comment pair has common keywords ‘movement’ and ‘GK’, and the temporal durations’ relationship of the two keyword pairs is the same, ‘finishes’.

This system extracts scenes of the keeper in a strange movement in the shot event. And the images of these scenes are similar. In this way, scene retrieval concerning object can extract similar image scenes only by analyzing comment information.

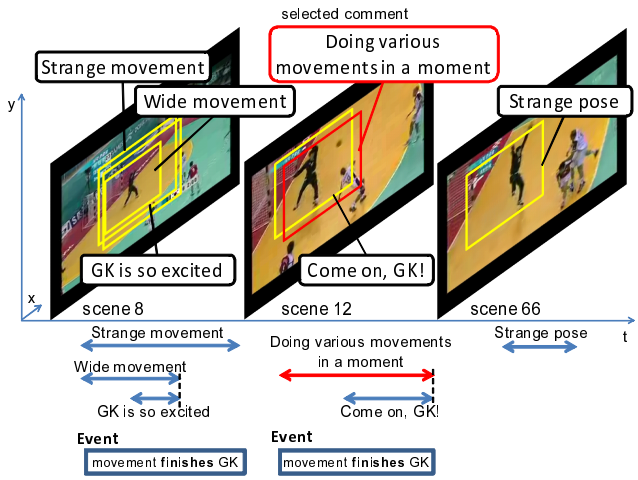

図 10　オブジェクトに関連して抽出されるシーンの例
Fig. 10　An example of scenes concerning object

5.3.2 Example of scenes related to event

We explain an example of scenes that are extracted as the result of the scene retrieval concerning event. In Fig. 11, when the user watches the shared video of the soccer game and selects the comment “The keeper’s reaction is good” that was written to the goalkeeper about the shot event in scene 21, the system extracts 2 other scenes as relevant scenes. Scene 1 is extracted because the scene relation is $EVENT$, or there are the comment pairs “The keeper’s reaction is so good” and “Nice save” or “Nice keep” that determined the same event of the selected comment and the other comment “Nice shot and keep” in scene 21. These comment pairs have common keywords “keeper” or “reaction” or “good” and “nice”, and the temporal durations’ relationship of two keyword pairs is the same, ‘meets’. Also scene 16 is extracted because the scene relation is $EVENT$, or there is the comment pair “The keeper isn’t moving” and “Stopped” that determined the same event of the selected comment and the other comment “Stopped” in scene 21. The comment pair has common keywords “keeper” and “stop”, and the temporal durations’ relationship of two keyword pairs is the same, ‘starts’.

The scene retrieval concerning event extracts scenes of the shoot event that users respond with a similar reaction. We consider that it especially suits videos of soccer, handball, basketball, etc. This is because sports tend to have the same event happening patterns; for example, the shooter kicks the ball, the keeper tries to stop the ball.

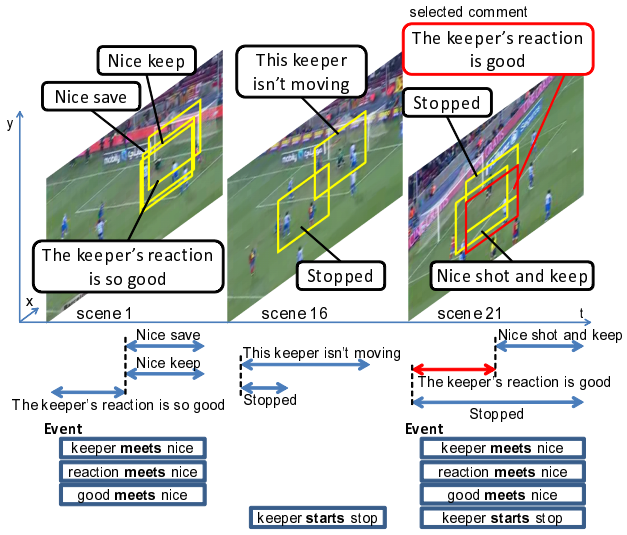

図 11　イベントに関連して抽出されるシーンの例

Fig. 11　An example of extracted scenes concerning event

6. Related works

Masuda et al. [9], [10] developed "Synvie", an application based on annotations acquired from the video blog community. In this system, the user can write comments with detailed information: object position, time range, object for comments, type of comment, comment, name, URL, and evaluation. In addition, the user can communicate with other users using a blog function. The user can create a blog entry about any video scene. Miyamori et al. [11] developed a system using on-line chat for a TV program. This system automatically extracts popular scenes on the basis of face marks and the number of written comments made for a particular scene. These systems aim to retrieve video scenes using annotated comments, blog entries, and on-line chatting by analyzing their text. On the other hand, Kitayama et al. [15] developed a method that generates comment sets using temporal duration and the pointing region for the purpose of organizing large numbers of comments in video sharing systems. Though they use temporal duration and the pointing region, text information is not considered. In our method, however, we aim to extract comments for viewing video by analyzing not only text information but also non-text information.

Fukuhara et al. [12] proposed a system for collecting and analyzing blog articles to gain an understanding of the concerns of people from collective and personal viewpoints. Their approach 1) analyzes relationships between blog articles and real temporal data, 2) extracts a topic of interest, and 3) identifies trends. Glance et al. [13] proposed a system called BlogPulse, which extracts trends from collected blog articles. Using keyword occurrence rates over a given period of time, the system classifies current trends. In communication via shared video, the user does not necessarily need to know what topic is being discussed and so our method simply extracts related comments.

Kimura et al. [14] developed an editing support system using user gaze of video. In this system, user gaze represents the user's viewing of temporal duration and the pointing region. Their approach is very similar to ours as the method extracts important scenes and regions using temporal duration and the pointing region. Pradhan et al. [16] proposed a method of glue joining that generates a new video interval from a keyword interval set. They use temporal duration for generating intervals, whereas we extract a related comment using temporal duration and the pointing region.

7. Concluding Remarks

We have proposed a method for extracting scenes based on not only the text information but also the non-text information of user comments, the pointing region and temporal duration. We aim at object and event that specified by the text and pointing region or temporal duration of comments. And comments that indicate the same object or event as the selected comment are determined by relations of pointing regions or temporal durations. Our proposed system retrieves scenes according to their relation to one another as defined by our set of relation types: $EQUAL$, $OBJECT$, $EVENT$ and $NO\ RELATION$. These are determined by the comments' reference to objects and events in video scenes. In the experiment in order to evaluate about the extracted scenes, our system was able to extract relevant scenes by using user comments, and these scenes cannot be extracted using the matching keywords method. The relevant scenes concerning object are similar as the user selected scenes' images. The relevant scenes concerning event are similar as the user selected scenes' situations.

In the future, we have to improve the interface in order to be able to write comments in an easier way. We also have to improve the determination method of both object and event in order to apply it to other types of videos and we have to find a way to classify users in order to prevent comments written by the same user from being extracted.

Acknowledgments

This research was supported in part by a Grant-in-Aid for Scientific Research (B)(2) 20300039 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

文　献

[1] YouTube. http://www.youtube.com/.
[2] Google Video. http://video.google.com/.
[3] nico nico douga. http://www.nicovideo.jp/.
[4] jimaku.in. http://jimaku.in/.

[5] Mecab. http://mecab.sourceforge.net/.
[6] SlothLib. http://www.dl.kuis.kyoto-u.ac.jp/slothlib/.
[7] L. Chaisorn, T.-S. Chua, and C.-H. Lee. Extracting Story Units in News Video. In Proc. of International Workshop on Advanced Image Technology 2003 (IWAIT 2003), 2003.
[8] I. Ide, K. Yamamoto, R. Hamada, and H. Tanaka. An automatic video indexing method based on shot classification. In Systems and Computers in Japan, volume 32, pages 32-41, August 2001.
[9] T. Masuda, D. Yamamoto, S. Ohira, and K. Nagao. Video Scene Retrieval Using Online Video Annotation. In Lecture Notes on Artificial Intelligence (LNAI 4914: JSAI 2007 (K.Satoh, et al. Ed.). Springer-Verlag, 2008.
[10] D. Yamamoto and K. Nagao. iVAS: Web-based Video Annotation System and its Applications. In Proc. of International Semantic Web Conference 2004, 2004.
[11] H. Miyamori, S. Nakamura, and K. Tanaka. Generation of Views of TV Content Using TV Viewers 'Perspectives Expressed in Live Chats on the Web. In Proc. of the 13th Annual ACM International Conference on Multimedia (ACM Multimedia2005), pages 853-861, 2005.
[12] T. Fukuhara, T. Murayama, and T. Nishida. Analyzing concerns of people using Weblog articles and real world temporal data. In Proc. of WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 2005.
[13] N. S. Glance, M. Hurst, and T. Tomokiyo. BlogPulse: Automated Trend Discovery for Weblogs. In Proc. of WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 2004.
[14] T. Kimura, H. Tanaka, and K. Sumiya. A Video Editing Support System using Users 'Gazes. In Proc. of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim 2005), pages 149-152, 2005.
[15] D. Kitayama, N. Oda, K. Sumiya. A Social Video Sharing System using User Comments based on Temporal Duration and Pointing Region. In Proc. of International Workshop on Information-explosion and Next Generation Search (INGS2008), pp.55-58, 2008.
[16] S. Pradhan, K. Tajima, and K. Tanaka. A Query Model forRetrieving Relevant Intervals within a Video Stream. In Proc. of the 6th IEEE Int l Conference on Multimedia Computing and Systems (ICMCS 99), volume 2, pages 788-792, 1999.
[17] J.F.Allen. Maintaining Knowledge about Temporal Intervals. In $Communications of the ACM, Vol.26$, pages 832-843, 1983.