



**HAL**  
open science

# Temporal relation algebra for audiovisual content analysis

Zein Al Abidin Ibrahim, Isabelle Ferrané, Philippe Joly

► **To cite this version:**

Zein Al Abidin Ibrahim, Isabelle Ferrané, Philippe Joly. Temporal relation algebra for audiovisual content analysis. *Multimedia Tools and Applications*, 2018, 78, pp.15275-15316. 10.1007/s11042-018-6771-1 . hal-02089343

**HAL Id: hal-02089343**

**<https://hal.science/hal-02089343v1>**

Submitted on 3 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:

<http://oatao.univ-toulouse.fr/22692>

### Official URL

DOI : <https://doi.org/10.1007/s11042-018-6771-1>

**To cite this version:** Ibrahim, Zein Al Abidin and Ferrané, Isabelle and Joly, Philippe *Temporal relation algebra for audiovisual content analysis*. (2018) *Journal of Multimedia Tools and Applications*, 78 (309). 1-42.  
ISSN 1380-7501

Any correspondence concerning this service should be sent to the repository administrator: [tech-oatao@listes-diff.inp-toulouse.fr](mailto:tech-oatao@listes-diff.inp-toulouse.fr)

# Temporal relation algebra for audiovisual content analysis

Zein Al Abidin Ibrahim<sup>1</sup>  · Isabelle Ferrane<sup>2</sup> · Philippe Joly<sup>2</sup>

## Abstract

The context of this work is to characterize the content and the structure of audiovisual documents by analysing the temporal relationships between basic events resulted from different segmentations of the same document. For this objective, we need to represent and reason about time. We propose a parametric representation of temporal relation between segments (points or intervals) in which the parameters are used to characterize the relationship between two non-convex intervals corresponding to two segmentations in the video analysis domain. The relationship is represented by a co-occurrences matrix noted as Temporal Relation Matrix (TRM). Each document is represented by a set of TRMs computed between each couple of segmentations of the same document using different features. The TRMs are analysed later to detect semantic events, highlight clues about the video content structure or to classify documents based on their types. For higher-level semantic events and documents' structure, we needed to apply some operations on the basic temporal relations and TRMs such as composition, disjunction, complement, intersection, etc. These operations brought to light more complex patterns; e.g. event 1 occurs at the same time of event 2 followed by event 3. In the work presented in this paper, we define a temporal relation algebra including its set of operations based on the parametric representation and TRM defined above. Several experiments have been done on different audio and video documents to show the efficiency of the proposed representation and the defined operations for audiovisual content analysing.

**Keywords** Audiovisual document analysis · Classification · Structuring · Representation · Event detection · Temporal relations algebra

✉ Zein Al Abidin Ibrahim  
zein.ibrahim@ul.edu.lb

Isabelle Ferrane  
Isabelle.Ferrane@irit.fr

Philippe Joly  
Philippe.Joly@irit.fr

<sup>1</sup> LARIFA Team, Faculty of Sciences – Hadath, Lebanese University, Beirut, Lebanon

<sup>2</sup> SAMOVA Team, IRIT, University of Paul Sabatier, Toulouse, France

## 1 Introduction

In the domain of audiovisual content indexing, several automatic tools have been already presented to help in segmenting a video stream and detecting specific objects in it. Most of these methods are based on low-level features extracted in a previous step (activity rate, dominant color, music or speech on the soundtrack, text or character presence on the screen, etc.). The goal of these techniques is to detect events, in order to create summaries or to determine temporal entries in the video stream as inputs for an index or for further processing steps. Their results generally suffer from the lack of semantics of produced index. The important information used during the analysis process might come from video, audio, or even textual sources. Thus, the need for tools that can combine the separated information to improve the semantic level of the extracted data became an issue. This could be considered as a first step towards more complex processes. An interesting way to detect relevant structuring rules, to extract its structural schema or to categorize the content type of an audiovisual document independently and without supervision is to study its temporal structure. For example, observing alternation between events of high semantic levels like newscaster on the screen and report will be more characteristic of a TV news program than a TV show.

Currently, indexing tools can be classified based on the features involved in the analysis process. In the literature, features extracted from motion analysis are used in [60] to detect the complete set of events that can be present in a soccer game video based on tracking players and ball in the field. Since the performance of such techniques is based on the accuracy of the players and ball tracking technique, background subtraction and colour features have been used in several works to improve the performance of tracking techniques such as in [26]. Motion features have been also used to extract highlights and events from soccer games [7]. Some other techniques are based on colour and movement features like in [70] to classify the soccer video game in two phases: play and break. Colour and shape are used as in [3] to classify the news video in semantic classes. The classification of shots in play and break was used in [59] to detect several types of semantic events like goal, card, goal attempt, corner, foul, offside, and non-highlights. The method extracts features from each play and break sequence and then applies Bayesian network to recognize the mentioned events. Classification of basketball videos in semantic classes using some rules is made as in [75] by a combination of colour, shape, movement, and texture. Other techniques base on audio features like in [2] aiming to detect audio events for highlights generation in which the rugby sport is used as a case study, or to generate summary for tennis sports video as in [57]. Multimodal features are also used in [18] to index the news videos by detecting 6 semantic classes, in [23] to detect the highlights of baseball videos and in [45] to extract the highlights in TV programs of formula 1.

These techniques have several limitations. One of them resides in the use of prior knowledge about analyzing the video data. This knowledge concerns what we are looking for, either what is the type of the video we are analyzing or the rules that exist in the domain the video belongs to (rules used in a soccer game, rules used in the production phase for this type of videos, etc.). In this case, the scope of these techniques is limited to a specific content and must be updated each time the knowledge changes. Hence, it cannot be used neither to analyze new types of content nor to retrieve events that are not predefined.

Some efforts to generalize event detection techniques are made. For example, in [16] generalization is made at sports video level but this is still limited to a specific domain.

With the development of video content understanding applications, and the success of deep learning techniques in image [20, 33, 39, 53, 54, 74] and speech domains [22, 25], several

works have been proposed for video content analysis using deep learning techniques, such as object recognition (Faster R-CNN [52], SSD [68], and YOLO ([50, 51])), video classification [31, 48, 49, 56, 72], video captioning [15, 71], congestion detection and crowd counting (i.e. [66]), scene labelling (i.e. [67]) and many other applications. Up to our day, the temporal information is not well taken into account. For example, image-based video classification methods such as [49] treat a video as a collection of independent frames where each is represented by a feature vector derived from a fully connected layer. Then the feature vector of the whole video is extracted as the average of all frames' feature vectors. The feature vector is fed later to any traditional algorithm such as SVM or Random Forest for recognition. The methods that tried to take into account temporal information stay limited to low level descriptors like optical flow or trajectories. Hence, the structure of videos is not fully represented. However, a possible benefit of deep learning methods is to feed the set of 3D co-occurrence matrices (the TRMs) into 3D CNN in order to train a model to recognize the structural information inside videos.

A promising proposition to cope with the above-mentioned limitations could be to rely on several segmentation systems that provide information on the content evolution at a low and mid-level of semantic. This would be done by observing the temporal relations between events coming from the resulted segmentations of different systems. Then, deducing some complementary descriptive information on the document temporal structure or content. Unlike the previously used techniques for video content analysis, the proposition here is not to use any predefined semantic segments or prior information. It is based on the mining of temporal segments provided by the different segmentation systems on different medias (image, audio).

To reach this goal, we came to the main contribution of the work presented in this paper which is to define new temporal relation algebra. This algebra is composed of a parametric representation of temporal relations between convex segments (a segment that has no gaps) and the set of well-known operations that exists in the literature such as composition, complement, disjunction etc. The proposed temporal model is a hybrid model that handles quantitative and qualitative temporal information. Moreover, the parametric representation of temporal relation can be calculated between points and intervals. In this algebra, the parametric representation is also extended to work for non-convex intervals (defined as the union of several ones) that are used to compute a co-occurrence matrix named Temporal Relation Matrix (TRM). Each audiovisual document is then represented by a set of TRMs. Each TRM in its turn is analyzed using the operations defined in the algebra, in order to highlight some frequent temporal patterns that exist between the two segmentations e.g. "when event 1 occurs, it is followed by event 2" or more complex ones such as "event 1 overlaps with event 2 that occurs at the same time with event 3". Since we represent each video by a set of TRMs that should be analyzed, the operations in the algebra are also defined to work on the TRMs. Later in this paper, we will show how many well-known temporal models can be derived from our proposed algebra. To test the efficiency of our contribution, several experimentations were done on different audio and video documents for video structuring, clustering and classification.

This paper will be organized as follows: in Section 2, we present the existing temporal relational algebras. Section 3 will be dedicated to our temporal model to represent temporal information. In this section, we present our parametric representation of temporal relations. We show how to compute what we call temporal relation matrix (TRM) that represents a relation between two non-convex intervals and we present a new reasoning method in Section 4. Section 5 will be dedicated to some notations that will be used through the article. In the

Section 6, we define the operations that can be applied on the new representation of temporal relations. In Section 7, we show how such operations can be applied between TRMs. Through Section 8, we validate the operations that we have defined on existing temporal models such as Allen's temporal relations. In Section 9, we show how operations can be applied on new relations derived from the distribution of votes in the TRMs. We end the section with the definition of a temporal relations algebra and then we show in Section 10 some already published works using this algebra to analyze audiovisual content. Finally, we conclude in Section 11.

## 2 Temporal representation and reasoning

Temporal representation and reasoning is an essential process in any activities that changes over time. That is why we find such process in several disciplines such as natural language processing, audio-visual content analysis, specification and verification of processes, temporal planning etc. The reader can find a list of possible applications in [21].

Hayes has introduced in [24] a basic representation of time in which six notions of time to represent temporal relations are given: basic physical dimension, time-line, time intervals, time points, amount of time or duration, and time positions. These notions were later used by several researchers in order to represent and to reason about time. Reader can find an overview of different approaches of temporal representation and reasoning in [9, 10], a survey in [62], and a review in [44]. However, we found that it is interesting to give a quick overview of the existing temporal representation and reasoning approaches in the literature.

Several temporal models to represent and reason about time were proposed in the literature. They can be classified according to the type of temporal entities they consider (point, interval, or both) or according to the type of temporal relations they deal with (qualitative, quantitative, or both). The qualitative models focus on the nature of the relations observed between the entities such as the relation before in the Allen's algebra. In contrast, quantitative models represent numerical values between the entities such as the distance between two entities, the duration of entities and so on.

In the literature, we could identify three well-known formalism that deal with qualitative temporal relations: Vilain and Kautz's Point Algebra [64] that handles temporal relations between points, Allen's Interval Algebra [1] that handles relations between intervals, Vilain's Point-Interval Algebra [63] and Ligozat's Generalized Interval Calculus [37] which are considered as hybrid models integrating the point and the interval entities in the same model.

In [64], the entities considered are time points and three basic temporal relations between points are defined: *before* ( $<$ ), *after* ( $>$ ) or *simultaneous* ( $=$ ). The temporal relation between two points may be a disjunction of the three basic relations if it cannot be defined. For example, we know that a point  $p_1$  is not before another point  $p_2$ . In this case, the relation between the two points  $p_1$  and  $p_2$  is the set  $\{=, >\}$ . That is why the three basic relations defined above represent only the cases when the two points are known. So, the relation that may exist between two points is a set of disjunctions  $\{\emptyset, \{<\}, \{>\}, \{=\}, \{<=\}, \{<,>\}, \{>=\}, \{<=,>\}\}$ . In multimedia systems, an example of a point-based representation is the timeline, on which media objects are placed on several time axes. Though this representation is also used as an interval-based representation, we can find the timeline model applied in various applications such as HyTime [30].

Allen introduced in [1] the famous and well-known interval algebra. The entities considered are intervals represented by their starting and ending times. The considered intervals are noted as convex ones in order to differentiate them from other type of intervals known as non-convex ones. In his algebra, Allen proposed a set of 13 temporal relations that may exist between two intervals:  $\{=, <, >, m, mi, o, oi, s, si, d, di, f, fi\}$ . Since an interval is represented by two points (its start and its end), a switch between the models is made by representing the intervals relations as conjunctions of point basic relations between the interval boundaries [40]. The Allen's algebra consists of the 8192 possible relations between intervals together with the operations inverse  $^{-1}$ , intersection  $\cap$ , and composition  $\wedge$ .

Since the computational complexity of Allen's formalism is intractable, several works in the literature tried to identify subclasses of the Allen's algebra that are tractable [43, 61, 65]. Beek et al. in [61] have defined the *pointisable algebra* as being the set of relations in the Allen's interval algebra that can be expressed by one of the relations  $<, \leq, =, \neq, \geq,$  and  $>$ . Vilain et al. defined in [65] the algebra of *Continuous Endpoint* (CEA) in which they model only continuous relations between time points. The algebra represents the set of the Allen's interval algebra which can be expressed by the  $<, \leq, =, \geq,$  and  $>$ . Nebel et al. in [43] have defined the ORD-Horn algebra basing on the notion of ORD clause. This clause is defined as the disjunction of relations having the form  $x R y$  where the relation  $R$  is one of the relations  $\leq, =,$  and  $\neq$ .

Some works focused on providing another representation of temporal relations or extending Allen's relations [38, 47, 73]. In [73], each interval  $I = [I_b, I_e]$  ( $I_b$  stands for interval beginning and  $I_e$  for interval end) is represented by the five zones:  $]-\infty, I_b[$ ,  $\{I_b\}$ ,  $]I_b, I_e[$ ,  $\{I_e\}$ , and  $]I_e, +\infty[$ . Using this representation, each Allen's temporal relation is represented by a  $5 \times 5$  matrix in which each value indicates the intersection between the associated zones. Contrary to the previous representation, Pujari et al. have extended the set of Allen's relations by integration of the duration information [47]. In this representation, each Allen's relation is superscripted by one of the relations  $\{<, =, >\}$  to express the new information about duration. For example, the meet relation noted as  $m$  becomes  $\{m^<, m^=, m^>\}$ . Ligozat et al. [38] have provided a graphical representation of the Allen's relations by regions. Each temporal relation is associated to a region in the Euclidean space.

Unbounded intervals are considered by Cukierman et al. who extend Allen's temporal relations to work with unbounded intervals [12]. The considered unbounded intervals are *since interval* with a finite beginning point and an infinite ending point, *until interval* with an infinite beginning point and a finite ending point, and *alltime* representing the time line with both infinite boundaries. Another work handling incomplete information about the start or the end of intervals was the one defined by Freksa in [19].

In the interval algebra of Allen, intervals are considered as convex ones which are intervals with no gaps. Ladkin defines in [35, 36] the notion of non-convex intervals defined as the union of convex ones. In this work, the defined temporal relations are based on the qualifiers *mostly*, *always*, *partially* and *sometimes*, and a *disjunction* relation to represent relation alternatives. The algebra defined in [35] has the advantage of being independent of the number of subintervals of each non-convex interval (potentially indefinite). It generates non-convex relations from convex ones.

Hybrid qualitative models have taken considerable place in the literature [37, 40, 63]. Vilain presented in his work [63] the temporal relations that may exist between a point and an interval and between an interval and a point. It allows the temporal qualitative relations between

objects of different types. In [17], a set of models that base on the temporal relations between intervals and points to compose multimedia data is cited (i.e. [8, 14]).

By the same way, Meri proposed in [40] a qualitative algebra in which a qualitative constraint between two events  $e_i$  and  $e_j$  (each may be a point or an interval), is a disjunction of the form:  $(e_i R_1 e_j) \vee (e_i R_2 e_j) \vee \dots \vee (e_i R_k e_j)$ , where each of the  $R$ 's is a basic relation that may exist between two objects. From this representation, the interval-interval relations, the point-point relations, the point-interval, and interval-point relations can be deduced.

Based on the previously presented formalisms, Ligozat proposed a generic notion of points, intervals and relations between them dealing with convex and non-convex intervals and points [37]. The proposed framework is based on the Vilain's point-interval relations [63] and Ladkin's non-convex interval ones [35]. In this approach, an interval is defined as a linearly ordered sequence of distinct points where a sequence of  $p$  points is called a  $p$ -interval. Consequently, a point is represented by a 1-interval while Allen's interval is a 2-interval. A 3-interval may represent three points, a point followed by an interval, or an interval followed by a point. Relations between a  $p$ -interval and a  $q$ -interval are called  $(p,q)$ -relations and noted by  $\Pi(p,q)$ .

Quantitative models are those which focus on quantitative relations rather than qualitative ones. An example of such model is the point-based distance algebra DA proposed by Dechter et al. in [13]. The DA allows to represent quantitative information between entities. It models distances between time points, durations of intervals, and allows constraints about the value of dates.

Some efforts have been dedicated to the proposition of temporal models integrating qualitative and quantitative information in the same framework [32, 40]. In [32], Kautz et al. have augmented the Allen's algebra with quantitative constraints of the form  $-c R_1 (x-y) R_2 d$  where  $R_1$  and  $R_2 \in \{<, \leq\}$  and  $x, y$  are the endpoints of the intervals. In the Meiri's temporal model [40], four types of qualitative constraints are taken into account: constraints between two points, constraints between a point and an interval, constraints between an interval and a point, and constraints between two intervals as already presented. The quantitative information is similar to the one presented in the DA by Dechter et al. [13].

To reason about time, several reasoning mechanisms are proposed in the literature. One of the well-known reasoning mechanisms is to consider relation between temporal entities as temporal constraints which are represented by a temporal network (i.e. [1]). The nodes of the network are the entities (point or intervals) and the vertices are the temporal relation that exist between the connected entities. The network is a special case of CSP (Constraint Satisfaction problems). Another particular CSP called Temporal CSP is used to represent quantitative information such as in [13] or qualitative and quantitative one such as in [32, 40]. A third type of network called the point-duration network (PDN) is used to reason about durations [11, 42, 46, 69]. The reader may refer to [55] for a survey of the constraint satisfaction problem (CSP) algorithms while Krokhn et al. provide a complete classification of the computational complexity of the algorithms of satisfiability of the IA [34].

In the multimedia domain, an event can be produced by interaction of multimedia objects, thus analysing temporal relations between events in an audio-visual document is an important issue. Results of such an analysis can be used to match up a given content with a predefined temporal structure (by means of hierarchical hidden markov models for example) in order to identify specific highlights, or to automatically build a temporal representation of the content evolution. The state of the art demonstrates that such tools are always built on a priori knowledge of how events are temporally related to each other in audio-visual documents.



For example, we can use the fact that anchor frames alternate with reports in a TV news program. By the same way, we can take into account that songs are followed by applause on entertainment program soundtracks, or that goal in a soccer game, could have been marked when the ball crosses the goal region and when an explosion of audiences' voice follows immediately. In the following section, we will introduce our generic parametric temporal model. Then, we will define our reasoning method having as a first aim the analysis of video content in order to detect information about the content and the structure of audio-visual documents.

### **3 Proposed temporal model**

The aim of our work is the multimedia content analysis especially events detection, video structuring, and audiovisual document categorization. These tasks are solved by analyzing the interaction between multimodal events. For example, a goal event in a soccer game is modeled as low movement segment followed by close up view of the players during an explosion on the audio track. By the same way, one of the ways to structure a news video is to identify the segments containing the newscaster and the segments that represent the reports in order to get the structure as a sequence of Newscaster segments (intervals) followed by Report ones. The interaction between multimodal events can be seen as the identification of temporal relations between segments of different types. The presence of frequent or rare temporal relations between segments may give valuable information about the structure and the content of the video document, as we will see later in this work.

In the video analysis domain, the temporal video segmentation task plays an essential role. This task is the process of partitioning video into temporal units that are homogeneous in some feature space. It is an important step of many video analysis problems such as video summarization, indexing, and retrieval. Different features and homogeneity criterion lead to different temporal segmentations. Each of the segmentations provides the temporal segments where this feature is homogeneous. Each segment (noted also event) may be represented by a temporal interval or point. Thus, temporal segment (event) is an important notion in the video analysis domain. Meanwhile, the analysis steps are applied on the audio track as the visual one, which leads to segments of different nature. In other words, audio segments have some duration while video segments may concern a frame sequence or one frame. Consequently, and based on the frame as basic unit, the segments may have duration and may not have. Thus, the mixture of the point-based and interval-based formalisms should be adopted. In other words, the best model is the one that handles interval-interval, interval-point, point-interval, and point-point temporal relations.

The exploitation of temporal relations between events may be qualitative-based, quantitative-based, or the mixture of the two. In our work, qualitative temporal relations may be seen as the target temporal relations to be considered. This is unfortunately false. In our framework, an event that lasts for 10 s should not be treated similarly to another lasting for 1 s. For example, an applause segment lasting for 10 s has not the same signification as a one that lasts for 1 s. Besides that, an event that occurs 1 min before another one may be semantically less related than two events with 10 s of distance. Other quantitative information may be also very valuable in the analysis step such as the intersection of two segments, time shifts between starts, ends, start and ends and so on. For these reason, the chosen model should consider not only qualitative information but also quantitative one about the temporal relations between

events in order to detect events of higher level. We should emphasize here that the aim of our work is not to compare our temporal model to the existing ones but to show that our model is an hybrid one and the existing models can be derived from our model.

The properties of the model that we should consider led us to propose a parametric representation of temporal relations between events. In the next section, we present our novel qualitative and quantitative representation handling temporal relations between segments of different types (points or intervals). Meanwhile, we show how existing temporal relations (Allen's interval-interval relations, Vilain's point-interval and interval-point relations, Vilain and Kautz's point-point relations...) can be derived directly from our parametric representation or derive new ones.

### 3.1 Parametric representation of temporal relation

Let  $I = [I_b, I_e]$  and  $J = [J_b, J_e]$  two temporal intervals characterized by their beginning and end times. The temporal relation that may exist between  $I$  and  $J$  is represented by the three parameters  $DE$ ,  $DB$ ,  $LAP$ . This representation is derived from the work of Moulin [41] used in the domain of natural language analysis. It allows us taking into account quantitative constraints and helps us to derive qualitative and quantitative constraints. They measure the time-shift between the boundaries of the two intervals. These parameters computed between  $I$  and  $J$  are presented in Fig. 1 and are defined as follows:

$$DE = J_e - I_e; DB = I_b - J_b; Lap = J_b - I_e$$

In the rest of this article, the relation between two intervals  $I$  and  $J$  is noted  $IR(DE, DB, Lap) J$ .

However, the parametric representation can be used to handle relations between a point and an interval. In this case, the point is represented by an interval with no duration ( $I_b = I_e$ ). Figures 2 and 3 show such representation.

As we can notice in the previous figures, two parameters are sufficient to represent the temporal relation between a point and an interval (and vice versa). In the interval-point representation, we have  $Lap = DE$  while in the point-interval one, we have  $DB = -Lap$ . Meanwhile, in the point-point parametric representation, we obtain  $DE = -DB = Lap$  which means that one parameter is sufficient to represent the relation between two points.

Using our parametric representation, we are able to deal with intervals or points and we can deduce directly if the temporal relation is an interval-interval, interval-point, point-interval, or point-point one based on the values of the parameters.

This parametric representation is mapped to a geometric one based on the three parameters. The computed parameters can be seen as the coordinate of a point in a 3D space. In other words, each temporal relation between two specific intervals (or points) can be represented

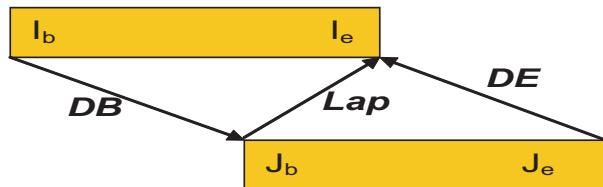


Fig. 1 Parameters of the temporal relation between two intervals

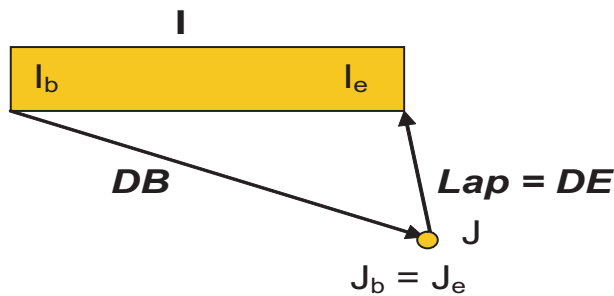


Fig. 2 Parametric representation of the interval-point relation

geometrically by a point in the 3D space. Figure 4 shows an example of three temporal relations computed between the following couple of intervals  $[I_1, J_1]$ ,  $[I_2, J_2]$ , and  $[I_3, J_3]$ .

Based on the parametric and geometric representations, the first question that should be asked is about the temporal relations that we may observe.

### 3.2 Which relations to observe?

The proposed parametric representation of temporal relations between events (points or intervals) can handle qualitative and quantitative information. There are two categories of temporal relations that can be observed. The first category represents predefined temporal relations such as the ones that have been proposed in the existing temporal models (Allen's relations...). The second category represents new relations that may be derived from the distribution of temporal relations between events. In the following paragraph, we show what type of relations can be observed in the first category. The second category will be presented later in the paper.

#### 3.2.1 Qualitative temporal information

**Relations between two points** We have previously presented the Vilain and Kautz's algebra that models the temporal relations that may exist between two points. Considering our parametric representation, the parameters are reduced to the following case:  $DE = -DB = Lap = J_e - I_e$  with  $J_b = J_e$  and  $I_b = I_e$ . Table 1 shows the different temporal relations between two points.

Relations between points can be represented graphically using a one-dimension space where the DE parameter is the value of the coordinate as presented in Fig. 5.

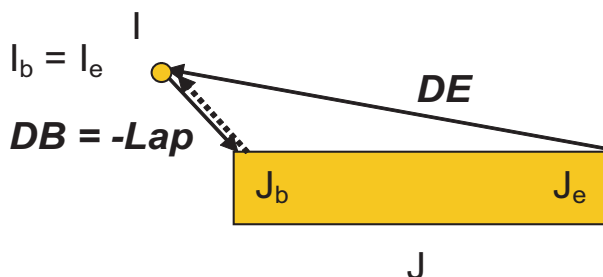


Fig. 3 Parametric representation of the point-interval relation

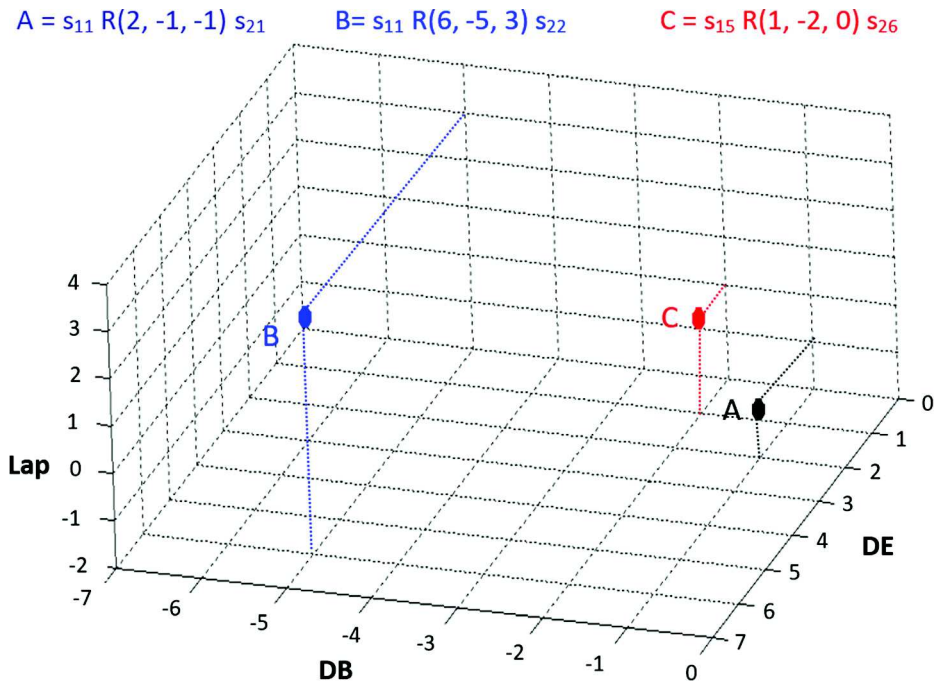


Fig. 4 Geometric representation of three temporal relations between intervals

**Relations between a point and an interval** In the Vilain's temporal model, a point may be related to an interval by one of the five following relations  $\{<, s, d, f, >\}$  as presented in the previous section.

Applying our parametric representation, we obtain  $DE = Lap$  which means that DE and DB are sufficient to represent the temporal relations between a point and an interval. The mapping between our representation and the five relations are presented in Table 2.

By the same way, the interval-point relations can be deduced directly from the DE, and DB parameters. Figure 6 shows the graphical representation of the temporal relations between a point and an interval.

As we can notice, the area corresponding to  $(DE < 0 \text{ and } DB < 0)$  is not associated to any temporal relation. This area represents the case where the start of an interval appears after its end (i.e.  $I_c < I_b$ ). This case is not considered here but such constraint is present in the literature. It models acyclic intervals [4, 5].

**Relations between two intervals** Based on the constraints between the boundaries of two intervals and by mapping these constraints to our representation space, we obtain a parametric representation of Allen's temporal relations. An example of such representation and how it is

Table 1 Parametric representation of the point-point relations

Relation	Constraint on DE
before	$DE > 0$
simultaneous	$DE = 0$
after	$DE < 0$



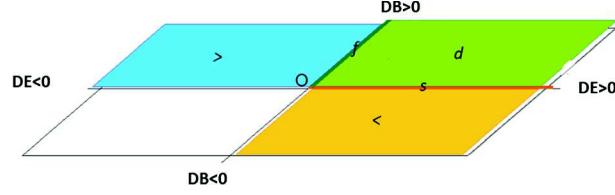


Fig. 6 Graphical representation of the point-interval temporal relations

Moreover, we can compute the distance noted as TShift separating two events:

$$TShift(I, J) = \begin{cases} Lap & \text{if } (Lap \geq 0) \\ Lap - DE + DB & \text{if } (Lap - DE + DB \geq 0) \\ Undefined & \text{otherwise} \end{cases}$$

The value of TShift(I,J) is set to “undefined” when the two events intersect.

The intersection quantity may be also deduced from the three parameters as follows:

$$\cap(I, J) = \begin{cases} 0 & \text{if } Lap \geq 0 \\ -Lap & \text{if } (Lap < 0, DB \leq 0, DE \geq 0) \\ DE - Lap - DB & \text{if } (Lap < 0, DB \geq 0, DE \leq 0) \\ \|I\| & \text{if } (Lap < 0, DB \geq 0, DE \geq 0) \\ \|J\| & \text{if } (Lap < 0, DB \leq 0, DE \leq 0) \end{cases}$$

By the same way, we can compute the union quantity of two events I and J as follows:

$$\cup(I, J) = \|I\| + \|J\| - \cap(I, J)$$

Table 3 Parametric representation of the interval-interval relations

Allen's relations	DE	DB	Lap
<	DE > Lap	DB < -Lap	Lap > 0
m	DE > 0	DB < 0	Lap = 0
o	DE > 0	DB < 0	Lap < 0
s	DE > 0	DB = 0	Lap < 0
f	DE = 0	DB > 0	Lap < 0
=	DE = 0	DB = 0	Lap < 0
d	DE > 0	DB < -Lap	Lap < 0
>	DE < 0	DB > 0	Lap > DE - DB
mi	DE < 0	DB > 0	Lap = DE - DB
oi	DE < 0	DB > 0	Lap < DE - DB
si	DE < 0	DB = 0	Lap < DE
fi	DE = 0	DB < 0	Lap < 0
di	DE < 0	DB < 0	Lap < DE

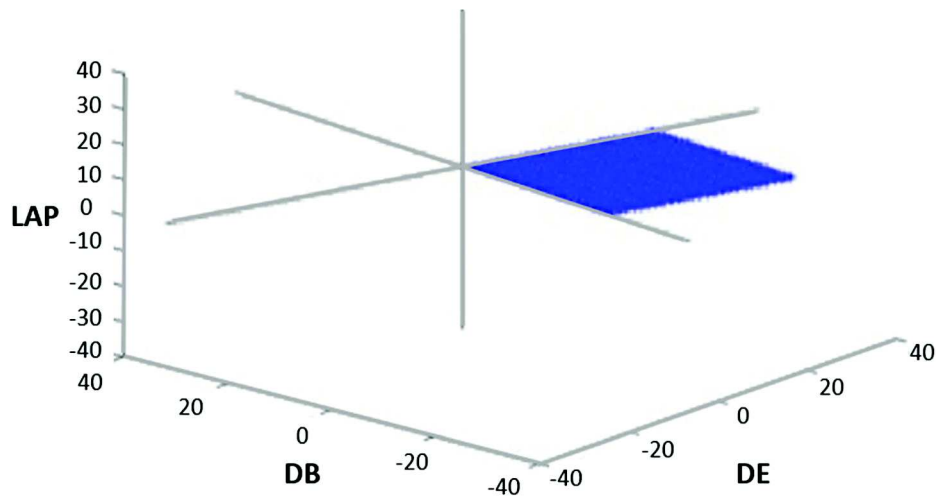


Fig. 7 Graphical representation of the interval-interval meet relation

#### 4 Matrix representation of temporal relations

A temporal relation between two segments can be represented using the three parameters mentioned above. Thus, from a graphical point of view, a relation between two intervals will be modeled as a 3D point. As we have already mentioned, in the video analysis domain, low-level features are mined in order to highlight some meaningful events in the content. In almost all the existing temporal segmentation methods, the start and the end of segments are usually produced based on one type of events. For example, one temporal segmentation system may localize all the segments of gradual transition effects, all appearances of a given character, moments where some music can be heard on the soundtrack, frames in which the same person appears, etc). The result of such segmentation contains meaningful information about the content of the video. Moreover, the events that belongs to different segmentations are not independent. For example, the appearance of a specific person on the screen may be

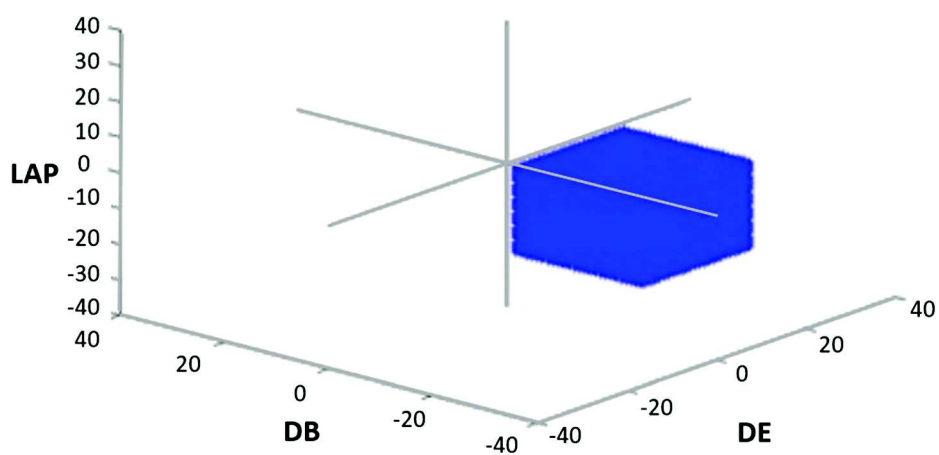


Fig. 8 Graphical representation of the interval-interval overlap relation

temporally related to the hearing of a music on the soundtrack. Thus, trying to highlight some temporal relations patterns is of big interest.

Let us consider that two temporal segmentations  $S_1$  and  $S_2$  of a same video document are performed in order to use their results in an analysis step. The segmentation corresponds to a set of temporally disjointed segments:

$$S_1 = \{s_{1i}\}_{i \in [1, M]} \quad S_2 = \{s_{2j}\}_{j \in [1, N]}$$

The segmentation can be viewed as a non-convex interval.

For two segmentations  $S_1$  and  $S_2$ , the three parameters between each couple of segments  $(s_{1i}, s_{2j})$  is evaluated and so represented by a point in the corresponding 3D space. The relation between  $s_{1i}$  and  $s_{2j}$  is noted as:

$$s_{1i} \mathbf{R}(DE, DB, LAP) s_{2j}$$

No existing temporal models that have been proposed in the literature to work with non-convex intervals are suitable in our context of work since the same model does not integrate quantitative and qualitative information and does not handle relation between different type of events (points and intervals). Hence, we propose to evaluate the temporal relation between two segmentations  $S_1$  and  $S_2$  (two non-convex intervals) by evaluating the temporal relation between each couple of intervals  $(s_{1i}, s_{2j})$ .

For each 3D point (ie. for each potential temporal relation between two intervals), we associate a vote accumulator that counts the number of times this relation is observed between the two segmentations (non-convex intervals). Then we obtain a matrix of accumulators called the Temporal Relation Matrix (TRM). It can be used directly to determine the frequencies of potential relations. It can also be used to observe remarkable distributions of votes and so to identify a general rule about the temporal behavior of events. For each occurrence of a given relation  $R$  (represented by the value of the vector of parameters), a vote is added to the associated cell in the TRM. For example, the relation  $s_{1i} \mathbf{R}(DE, DB, LAP) s_{2j}$  will correspond to the cell  $TRM[DE][DB][LAP]$ .

Figure 9 shows how the TRM is evaluated between the two segmentations  $S_1$  and  $S_2$  while Fig. 10 shows a graphical representation of the distribution of votes in a TRM.

Once the vote step has been performed, i.e. when all possible couples of segments have voted for the way they are temporally related, we have to analyze the TRM to identify for example the most frequent or rare relations between the concerned features. Unlike other vote techniques, a maximum value in a cell is not enough to fully identify a relation. Actually, most of semantic temporal relations determine subparts of the TRM where votes are distributed. Therefore, the first step of the TRM analysis is to localize zones in the 3D space regarding to the vote distribution. This localization can be achieved by clustering methods or classification methods. The Fig. 11 shows the graphical representation of a TRM computed between two segmentations in which three classes of relations are identified using the K-means clustering method. The temporal relation between the two segmentation is represented by a vector of three values that are the number of points in each cluster.

Another approach consists in using prior knowledge about semantic relations like for example the Allen's relations as presented in Section 2. This approach consists in identifying disjointed subparts of the vote space associated to remarkable relations. For example, the *overlap* relation in the Allen's algebra is associated to the 3D subpart  $overlap(X > 0, Y < 0, Z < 0)$ .



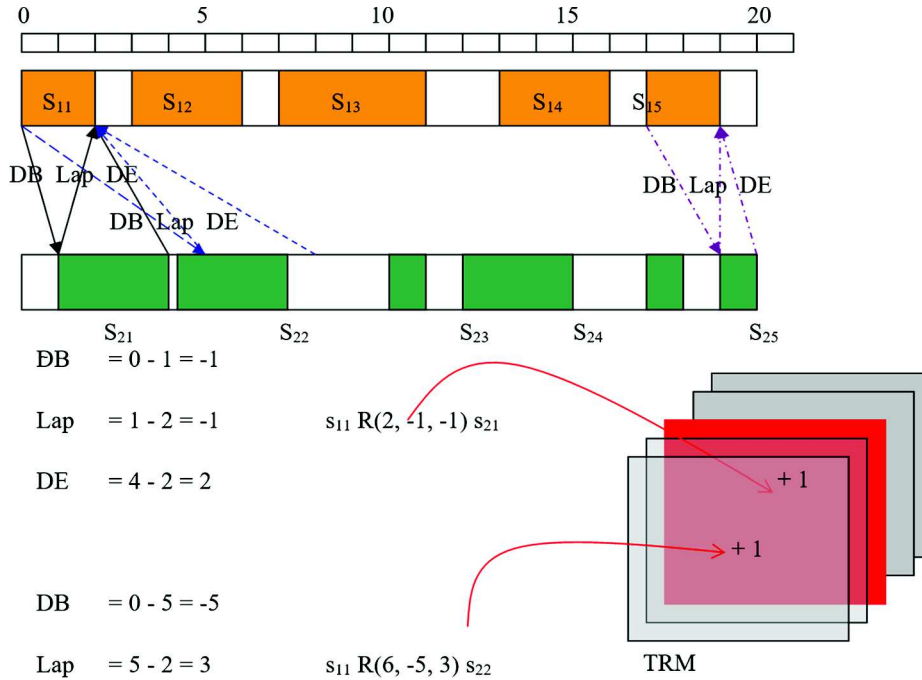


Fig. 9 Evaluation of temporal relations between two segmentations (non-convex intervals)

## 5 Notations

Before presenting the operations that we can apply on temporal relations, we present the notations that we use in to help you understand the rest of the article.

Let  $\mathcal{R} = \{(x, y, z) / (x, y, z) \in \mathbb{R}^3\}$  be the set of all possible relations that may exist between two events (points or intervals). We note  $S = P(\mathcal{R})$  the set of all possible subsets of  $\mathcal{R}$ .

In the rest of the article, we use the following notations:

- $R(A, B, C)$  in  $S$  where  $A$ ,  $B$  and  $C$  are convex intervals in  $\mathbb{R}$  (parameters in capital) represents a set of temporal relations and will be named temporal relations class. For example,  $R([-5, 10], [2, 9], [-7, 3])$  is a class of temporal relations.
- $R(a, b, c)$  (parameters in lower case) is a temporal relation between two specific intervals. For example,  $R(2, 7, 0)$  is a temporal relation.
- $R(a, b, c)$  is an instance of  $R(A, B, C)$  if and only if  $a \in A$ ,  $b \in B$ ,  $c \in C$ . In other words, we say that  $R(a, b, c)$  verify the constraints of  $R(A, B, C)$ . For example,  $R(2, 7, 0)$  is an instance of  $R([-5, 10], [2, 9], [-7, 3])$ .
- For each  $R(a, b, c)$  there exists two events  $I$  and  $J$  (intervals or points) that can be related by  $R$  such that  $I R(a, b, c) J$ .
- We note  $I R(A, B, C) J$  to express that  $I$  and  $J$  are related through a relation instance of  $R(A, B, C)$ .

## 6 Operations on temporal relations

In the literature, five operations on temporal relations are commonly used:

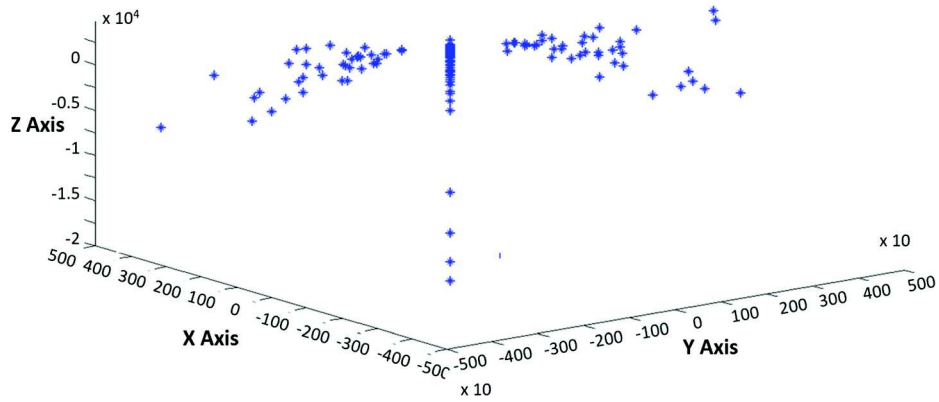


Fig. 10 Graphical representation of a TRM

- Disjunction or union: Undefinite relation between events is represented by the disjunction of possible relations that can present between two events I and J (intervals or points). The disjunction operators will be noted by the symbol  $\vee$ .

$$I (R \vee S) J \Leftrightarrow I R J \text{ or } I S J$$

- Composition: Let  $R_1$  (respect.  $R_2$ ) represents the relation between the events I and J (respect. J and K). The composition will be noted as  $R_1 \circ R_2$  and represents the relation that presents between I and K.

$$I (R \circ S) J \Leftrightarrow \exists K / I R K \ \& \ K S J$$

- Inverse: The inverse of the relation  $R$  holding between I and J is noted as  $R^{-1}$  and represents the relation between J and I.

$$I R^{-1} J \Leftrightarrow J R I \quad \forall (I, J)$$

- Intersection: The intersection of two classes of relations  $R_1(A_1, B_1, C_1)$  and  $R_2(A_2, B_2, C_2)$  represents the set of relations (the zone) where each relation verifies the constraints of  $R_1$  and  $R_2$ .

$$I (R \cap S) \Leftrightarrow I R J \ \& \ I S J$$

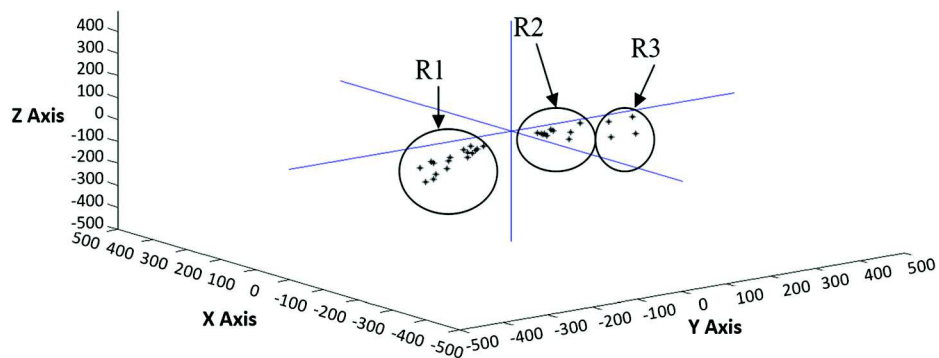


Fig. 11 TRM between two segmentations (two non-convex intervals)

- Complement: The complement of an interval relation  $R$  noted as  $C_R$  or  $R^c$  is the set of the remaining relations.

$$I R^c J \Leftrightarrow \neg(I R J)$$

## 6.1 Disjunction

The disjunction operator is a function that maps an element from  $S \times S$  domain to an element in the  $S$  domain.

$$f : S \times S \rightarrow S$$

$$(R_1(A_1, B_1, C_1), R_2(A_2, B_2, C_2)) \rightarrow f(R_1 \vee R_2) = R_1(A_1, B_1, C_1) \cup R_2(A_2, B_2, C_2)$$

In other words, the disjunction of two classes of relations represented by two subsets of  $Z^3$  is the union of these two subsets. For example, given the following two classes of relations  $R_1([7 \ 15], [20 \ 55], [-10 \ 15])$  and  $R_2([-15 \ -5], [-5 \ -5], [1 \ 25])$ , their disjunction is the union of the two subsets in the 3D space as shown in Fig. 12.

If each of the two classes of relations is reduced to a relation, the disjunction is the union of two 3D points.

## 6.2 Intersection

The intersection operator is a function that maps an element from  $S \times S$  domain to an element in the  $S$  domain.

$$f : S \times S \rightarrow S$$

$$(R_1(A_1, B_1, C_1), R_2(A_2, B_2, C_2)) \rightarrow f(R_1 \cap R_2) = R_1(A_1, B_1, C_1) \cap R_2(A_2, B_2, C_2) = R(A_1 \cap A_2, B_1 \cap B_2, C_1 \cap C_2)$$

In other words, the intersection of two classes of relations corresponds to the intersection of two 3D zone. For example, consider the following two classes of relations  $R_1([-5 \ 10], [1 \ 30])$ ,

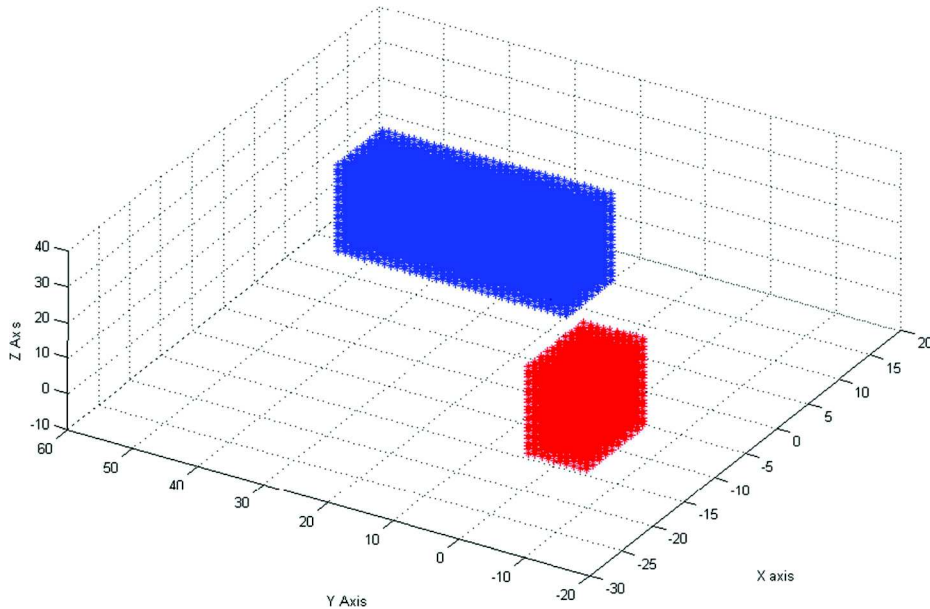


Fig. 12 Disjunction of  $R_1$  and  $R_2$

$[3\ 25]$ ) and  $R_2([-20\ -1], [-5\ 13], [1\ 5])$ . The intersection of  $R_1$  and  $R_2$  is the relation  $R([-5\ -1], [1\ 13], [3\ 5])$  as shown in Fig. 13.

### 6.3 Complement

The complement operator is a function that maps an element from the  $S$  domain to an element in the  $S$  domain.

$$f : S \rightarrow S$$

$$R(A, B, C) \rightarrow f(R) = R^c = R_1(\mathbb{R}-A, \mathbb{R}, \mathbb{R}) \vee R_2(A, \mathbb{R}-B, \mathbb{R}) \vee R_3(A, B, \mathbb{R}-C)$$

In other words, the complement of a class of relations  $R$  is equal to all the relations in  $S$  except the ones inside the zone of the class of relations  $R$ . For example, given the following class of relations  $R([-5\ 3], [10\ 25], [-6\ 2])$ , the complement of  $R$  in the space  $SPACE([-10\ 20], [-7\ 25], [-15\ 15])$  is shown on Fig. 14. Here, we have done the complement over  $SPACE$  and not over  $\mathbb{R}$  for plotting issues.

### 6.4 Composition

The composition operator is a function that maps an element from the  $S \times S$  domain into an element in the  $S$  domain.

$$f : S \times S \rightarrow S$$

$$(R_1(A_1, B_1, C_1), R_2(A_2, B_2, C_2)) \rightarrow f(R_1 \wedge R_2) = R(A, B, C)$$

$$= R(A_1 + A_2, B_1 + B_2, C_1 - B_2) \cap R^c(A_1 + A_2, B_1 + B_2, A_1 + C_2)$$

$$= R(A_1 + A_2, B_1 + B_2, (C_1 - B_2) \cap (A_1 + C_2))$$

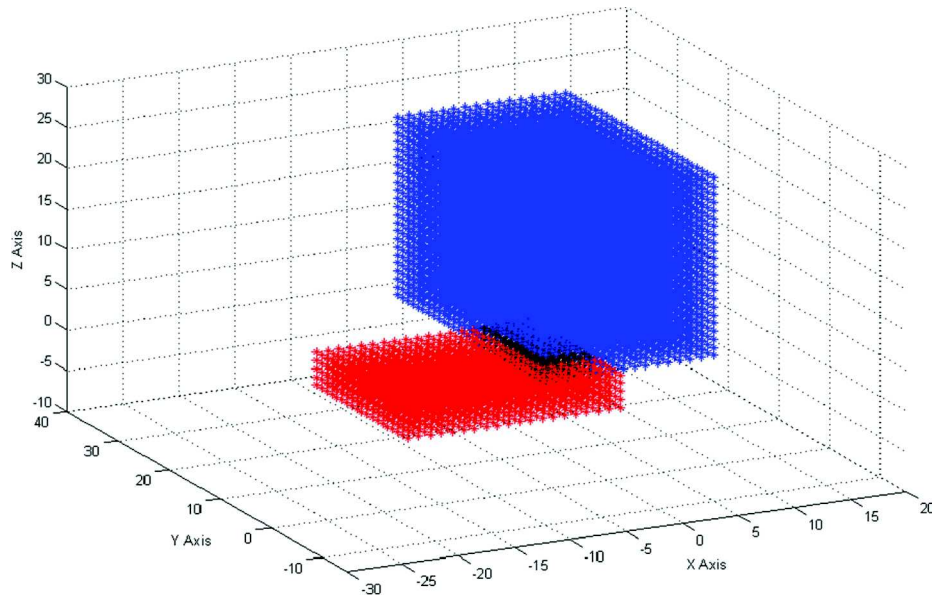


Fig. 13 Intersection of  $R_1$  and  $R_2$

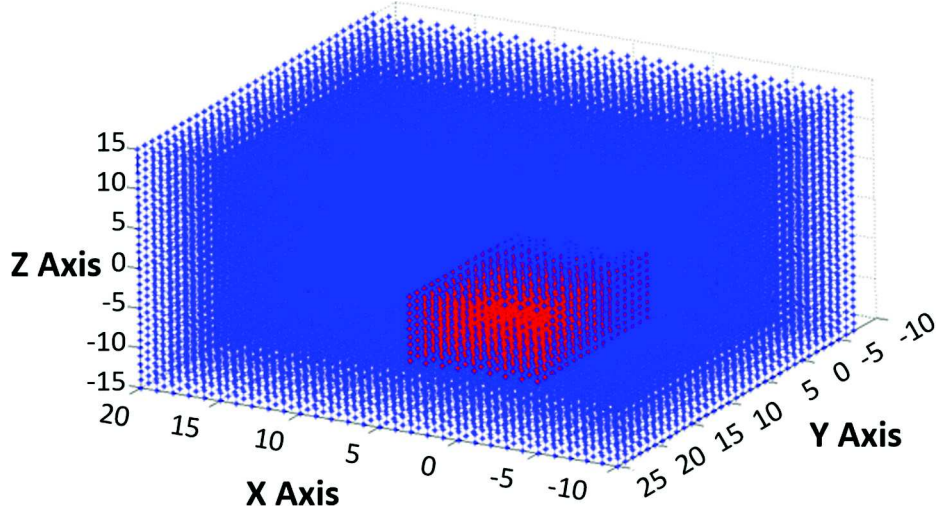


Fig. 14 Complement of R over SPACE

Before proving the above function, we start with the composition of two individual temporal relations instead of classes of relations.

$$f : S \times S \rightarrow S \\ (R_1(a_1, b_1, c_1), R_2(a_2, b_2, c_2)) \rightarrow f(R_1 \wedge R_2) = R(a, b, c) = R(a_1 + a_2, b_1 + b_2, c_1 - b_2) = R(a_1 + a_2, b_1 + b_2, a_1 + c_2)$$

**Proof** Let  $I = [I_b, I_e]$ ,  $J = [J_b, J_e]$ , and  $K = [K_b, K_e]$  three temporal intervals (sub-intervals) that correspond to three different segmentations (non-convex intervals).  $R_1(a_1, b_1, c_1)$ ,  $R_2(a_2, b_2, c_2)$  represent the relations between I and J, and J and K respectively. The composition of two relations  $R_1$  and  $R_2$  is defined by:

$R_1 \wedge R_2 = \{(I, K) \text{ where } I \text{ and } K \text{ are intervals} / \exists \text{ an interval } J, I R_1 J \text{ and } J R_2 K\}$ . Since the composition of the relations will represent the relation between I and K, in this case the parameters of the resulted relations are:

- $a = K_e - I_e = K_e - J_e + J_e - I_e = a_1 + a_2$ .
- $b = I_b - K_b = I_b - J_b + J_b - K_b = b_1 + b_2$ .
- $c = K_b - I_e = K_b - J_b + J_b - I_e = c_1 - b_2$ .
- $c = K_b - I_e = K_b - J_e + J_e - I_e = a_1 + c_2$ .

The idea behind the last equality is that the composition of two relations should have a common interval, which is in the above example the interval J.

To prove the composition operator for classes of relations, we proceed with proving the inclusion of each set in the other one. In other words, we should prove the following:

$$R_1(A_1, B_1, C_1) \wedge R_2(A_2, B_2, C_2) \subseteq R(A_1 + A_2, B_1 + B_2, (C_1 - B_2) \cap (A_1 + C_2))$$

and

$$R(A_1 + A_2, B_1 + B_2, (C_1 - B_2) \cap (A_1 + C_2)) \subseteq R_1(A_1, B_1, C_1) \wedge R_2(A_2, B_2, C_2)$$

**Proof** Let  $IR(a,b,c)J$  with  $I = [I_b I_e]$  and  $J = [J_b J_e]$ . Suppose that  $R$  is an instance of  $R_1(A_1, B_1, C_1) \wedge R_2(A_2, B_2, C_2)$ . We can write:

$$\begin{aligned}
I(R_1(A_1, B_1, C_1) \wedge R_2(A_2, B_2, C_2)) J &\Rightarrow \exists K = [K_b K_e] / I R_1(A_1, B_1, C_1) K K R_2(A_2, B_2, C_2) J \\
I R_1(A_1, B_1, C_1) K &\Rightarrow \exists (a_1, b_1, c_1) \in (A_1, B_1, C_1) / I R_1(a_1, b_1, c_1) K \\
K R_2(A_2, B_2, C_2) J &\Rightarrow \exists (a_2, b_2, c_2) \in (A_2, B_2, C_2) / K R_2(a_2, b_2, c_2) K \\
I R(a, b, c) J &\Rightarrow a = J_e - I_e = J_e - K_e + K_e - I_e = a_1 + a_2 \in (A_1 + A_2) \\
b = I_b - J_b &= I_b - K_b + K_b - J_b = b_1 + b_2 \in (B_1 + B_2) \\
c = J_b - I_e &= J_b - K_e + K_e - I_e = a_1 + c_2 \in (A_1 + C_2) \\
c = J_b - I_e &= J_b - K_b + K_b - I_e = c_1 - b_2 \in (C_1 - B_2)
\end{aligned}$$

Hence:

$$\begin{aligned}
R(a, b, c) \in R(A_1 + A_2, B_1 + B_2, C_1 - B_2) \text{ and } R(a, b, c) \in R(A_1 + A_2, B_1 + B_2, A_1 + C_2) &\Rightarrow R(a, b, c) \\
&\in R(A_1 + A_2, B_1 + B_2, (A_1 + C_2) \cap (C_1 - B_2))
\end{aligned}$$

Reciprocally,  $\forall I R(a, b, c) J \in R(A_1 + A_2, B_1 + B_2, C_1 - B_2) \cap R(A_1 + A_2, B_1 + B_2, A_1 + C_2)$ , we only have to find an interval  $K$  such that  $I R_1(A_1, B_1, C_1) K$  and  $K R_2(A_2, B_2, C_2) J$ .

$$\begin{aligned}
I R(a, b, c) J \in R(A_1 + A_2, B_1 + B_2, C_1 - B_2) \cap R(A_1 + A_2, B_1 + B_2, A_1 + C_2) &\Rightarrow a \in A_1 + A_2, b \in B_1 + B_2, c \in C_1 - B_2 \text{ and } c \in A_1 + C_2. \\
\Rightarrow \exists a_1, b_1, c_1, a_2, b_2, c_2 / a = a_1 + a_2, b = b_1 + b_2 \\
I R(a, b, c) J &\Rightarrow a = J_e - I_e, b = I_b - J_b, c = J_b - I_e.
\end{aligned}$$

By substituting the values of  $a$ ,  $b$ , and  $c$  in their places above and adding the values  $K_b$  and  $K_e$ , we can find the values of  $K = [K_b K_e]$  and hence we obtain  $I(R_1 \wedge R_2)J$ .

#### 6.4.1 Properties of the composition operator

The composition operator is a law of composition under the  $S$  domain ( $\langle S, \wedge \rangle$  is a law of composition). It has the following properties:

- $\langle S, \wedge \rangle$  is associative: Let  $R_1, R_2$ , and  $R_3$  be three temporal relations of  $S$  between the intervals  $I, J$  and  $K$ . we have:

$$R_1 \wedge (R_2 \wedge R_3) = (R_1 \wedge R_2) \wedge R_3$$

**Proof**

$$\begin{aligned}
R_1(A_1, B_1, C_1) \wedge ((R_2(A_2, B_2, C_2) \wedge R_3(A_3, B_3, C_3))) & \\
= R_1(A_1, B_1, C_1) \wedge R'(A_2 + A_3, B_2 + B_3, (C_2 - B_3) \cap (A_2 + C_3)) & \\
= R(A_1 + A_2 + A_3, B_1 + B_2 + B_3, (C_1 - B_2 - B_3) \cap (A_1 + C_2 - B_3)) & \\
\cap (A_1 + A_2 + C_3) & \\
(R_1(A_1, B_1, C_1) \wedge R_2(A_2, B_2, C_2)) \wedge R_3(A_3, B_3, C_3) & \\
= R'(A_1 + A_2, B_1 + B_2, (C_1 - B_2) \cap (A_1 + C_2)) \wedge R(A_3, B_3, C_3) & \\
= R(A_1 + A_2 + A_3, B_1 + B_2 + B_3, (C_1 - B_2 - B_3) \cap (A_1 + C_2 - B_3)) & \\
\cap (A_1 + A_2 + C_3) &
\end{aligned}$$

We can prove the associativity property by taking instance from each class of relation as follows:

Let  $R_1(a_1, b_1, c_1)$ ,  $R_2(a_2, b_2, c_2)$ ,  $R_3(a_3, b_3, c_3)$  be instances of  $R_1(A_1, B_1, C_1)$ ,  $R_2(A_2, B_2, C_2)$ ,  $R_3(A_3, B_3, C_3)$  respectively.

$$\begin{aligned} R_1(a_1, b_1, c_1) \wedge (R_2(a_2, b_2, c_2) \wedge R_3(a_3, b_3, c_3)) &= R_1(a_1, b_1, c_1) \wedge R'(a_2 + a_3, b_2 + b_3, c_2 - b_3) \\ &= R''(a_1 + a_2 + a_3, b_1 + b_2 + b_3, c_1 - b_2 - b_3) \end{aligned}$$

And

$$\begin{aligned} (R_1(a_1, b_1, c_1) \wedge R_2(a_2, b_2, c_2)) \wedge R_3(a_3, b_3, c_3) &= R'(a_1 + a_2, b_1 + b_2, c_1 - b_2) \wedge R_3(a_3, b_3, c_3) \\ &= R''(a_1 + a_2 + a_3, b_1 + b_2 + b_3, c_1 - b_2 - b_3) \end{aligned}$$

–  $\langle S, \wedge \rangle$  is not commutative:

$$R_1 \wedge R_2 \neq R_2 \wedge R_1$$

### Proof

$$\begin{aligned} R_1(A_1, B_1, C_1) \wedge R_2(A_2, B_2, C_2) &= R'(A_1 + A_2, B_1 + B_2, (C_1 - B_2) \cap (A_1 + C_2)) \\ R_2(A_2, B_2, C_2) \wedge R_1(A_1, B_1, C_1) &= R''(A_1 + A_2, B_1 + B_2, (C_2 - B_1) \cap (A_2 + C_1)) \end{aligned}$$

$\langle S, \wedge \rangle$  is commutative in the case where  $(C_1 - B_2) = (C_2 - B_1)$  and  $(A_1 + C_2) = (A_2 + C_1)$ . This case corresponds to intervals with equal duration. This can be simply proven by taking an instance of each relation as follows:

$$\begin{aligned} R_1(a_1, b_1, c_1) \wedge R_2(a_2, b_2, c_2) &= R'(a_1 + a_2, b_1 + b_2, c_1 - b_2) = R'(a_1 + a_2, b_1 + b_2, a_1 + c_2) \\ R_2(a_2, b_2, c_2) \wedge R_1(a_1, b_1, c_1) &= R''(a_1 + a_2, b_1 + b_2, c_2 - b_1) = R''(a_1 + a_2, b_1 + b_2, a_2 + c_1) \end{aligned}$$

$\langle S, \wedge \rangle$  is commutative if and only if:

$$\begin{aligned} c_1 - b_2 = c_2 - b_1 &\Leftrightarrow (J_b - I_e) - (I_b - K_b) = (K_b - J_e) - (I_b - J_b) \Rightarrow I_e - I_b = J_e - J_b \Rightarrow \text{duration}(I) \\ &= \text{duration}(J). \end{aligned}$$

If we take the case  $a_1 + c_2 = a_2 + c_1$ , we obtain the  $\text{duration}(J) = \text{duration}(K)$

In other words,  $\langle S, \wedge \rangle$  is commutative in the cases where I and J or J and K have the same duration. This result is logic since an event  $e_1$  that appears, for example, before a second event  $e_2$  is not the same as when  $e_2$  appears before  $e_1$ .

### 6.4.2 Properties of the disjunction operator

The disjunction operator is a law of composition under the S domain ( $\langle S, \vee \rangle$  is a law of composition). It has the following properties:

–  $\langle S, \vee \rangle$  is associative: Let  $R_1$ ,  $R_2$ , and  $R_3$  be three temporal relations of S between the intervals I, J and K. we have:

$$R_1 \vee (R_2 \vee R_3) = (R_1 \vee R_2) \vee R_3$$

**Proof**  $R_1(A_1, B_1, C_1) \vee ((R_2(A_2, B_2, C_2) \vee R_3(A_3, B_3, C_3))) = R_1(A_1, B_1, C_1) \cup ((R_2(A_2, B_2, C_2) \cup R_3(A_3, B_3, C_3)) = (R_1(A_1, B_1, C_1) \cup (R_2(A_2, B_2, C_2))) \cup R_3(A_3, B_3, C_3)$  since the union operator is associative over the  $\mathbb{R}^3$  space.

We can prove the associativity property by taking instance from each class of relation as follows:

Let  $R_1(a_1, b_1, c_1)$ ,  $R_2(a_2, b_2, c_2)$ ,  $R_3(a_3, b_3, c_3)$  be instances of  $R_1(A_1, B_1, C_1)$ ,  $R_2(A_2, B_2, C_2)$ ,  $R_3(A_3, B_3, C_3)$  respectively.

$$\begin{aligned} R_1(a_1, b_1, c_1) \vee (R_2(a_2, b_2, c_2) \vee R_3(a_3, b_3, c_3)) &= R_1(a_1, b_1, c_1) \vee R'(a_2 + a_3, b_2 + b_3, c_2 + c_3) \\ &= R''(a_1 + a_2 + a_3, b_1 + b_2 + b_3, c_1 + c_2 + c_3) \end{aligned}$$

And

$$\begin{aligned} (R_1(a_1, b_1, c_1) \vee R_2(a_2, b_2, c_2)) \vee R_3(a_3, b_3, c_3) &= R'(a_1 + a_2, b_1 + b_2, c_1 + c_2) \vee R_3(a_3, b_3, c_3) \\ &= R''(a_1 + a_2 + a_3, b_1 + b_2 + b_3, c_1 + c_2 + c_3) \end{aligned}$$

–  $\langle S, \vee \rangle$  is commutative:

$$R_1 \vee R_2 = R_2 \vee R_1$$

**Proof**

$$\begin{aligned} R_1(A_1, B_1, C_1) \vee R_2(A_2, B_2, C_2) &= R_1(A_1, B_1, C_1) \cup R_2(A_2, B_2, C_2) \\ &= R_2(A_2, B_2, C_2) \cup R_1(A_1, B_1, C_1) \\ &= R_2(A_2, B_2, C_2) \vee R_1(A_1, B_1, C_1) \end{aligned}$$

### 6.4.3 Other properties

– Identity relation: The class of relations  $R_e(\{0\}, \{0\}, \mathbb{R})$  is the identity class of relations of the composition operator. The following properties hold for each class of relations  $R(A, B, C)$ :

$$\begin{aligned} R(A, B, C) \wedge R_e(\{0\}, \{0\}, \mathbb{R}) &= R(A, B, C) \\ R_e(\{0\}, \{0\}, \mathbb{R}) \wedge R(A, B, C) &= R(A, B, C) \end{aligned}$$

More specifically, we have also the following properties that hold:

$$\begin{aligned} R(A, B, C) \wedge R_e(\{0\}, \{0\}, C-A) &= R(A, B, C) \\ R_e(\{0\}, \{0\}, B+C) \wedge R(A, B, C) &= R(A, B, C) \end{aligned}$$

**Proof**

$$\begin{aligned} R(A, B, C) \wedge R_e(\{0\}, \{0\}, \mathbb{R}) &= R(A, B, C - \{0\}) \cap R(A, B, A + \mathbb{R}) = R(A, B, C \cap \mathbb{R}) \\ &= R(A, B, C) \end{aligned}$$

By the same way,  $R_e(\{0\}, \{0\}, \mathbb{R}) \wedge R(A, B, C) = R(A, B, C)$ .



The above property can be proven for all instances of  $R(A,B,C)$ . Let  $R(a,b,c)$  be an instance of  $R(A,B,C)$ .

$$\exists R_e(0,0,b+c) \in R(\{0\}, \{0\}, B+C) \subseteq R(\{0\}, \{0\}, \mathbb{R}) / R(a,b,c) \wedge R_e(0,0,b+c) = R(a,b,c)$$

and

$$\exists R_e(0,0,c-a) \in R(\{0\}, \{0\}, C-A) \subseteq R(\{0\}, \{0\}, \mathbb{R}) / R_e(0,0,c-a) \wedge R(a,b,c) = R(a,b,c)$$

The relation  $R_e(0,0,b+c)$  is a right identity relation. This relation is a part of the equal relation. In the qualitative representation of temporal relations, they use it totally as an identity relation. In the cases where  $\langle S, \wedge \rangle$  is commutative,  $R_e(0,0,b+c)$  is also the left neutral relation. Otherwise,  $R_e(0,0,c-a)$  is the left identity relation. In other words, we can say that the two identity relations intersect in the zones where the intervals have the same durations.

The Figs. 15, 16, and 17 show three examples of a class of relations and its associated parts of the identity class.

- Inverse relation: For all relations  $R(a,b,c) \in R(A,B,C)$  in  $S$ , there exist  $R^{-1}(-a,-b,-a+b+c)$  in  $S$  such that  $R(a,b,c) \wedge R^{-1}(-a,-b,-a+b+c) = R'(a-a,b-b,c+b) = R_e(0,0,b+c) \in R_e(\{0\}, \{0\}, \mathbb{R})$  and  $R^{-1}(-a,-b,-a+b+c) \wedge R(a,b,c) = R_e(0,0,c-b) \in R_e(\{0\}, \{0\}, \mathbb{R})$ .

**Proof** Let  $R_1(a_1, b_1, c_1)$  be a relation between two intervals  $I = [I_b, I_c]$ , and  $J = [J_b, J_c]$  in  $S$  ( $I R_1(a_1, b_1, c_1) J$ ). Let  $R_2(a_2, b_2, c_2)$  be the inverse between  $I$  and  $J$  ( $J R_2(a_2, b_2, c_2) I$ ). We have the following values for the parameters:

- $a_2 = I_c - J_c = -a_1$ .
- $b_2 = J_b - I_b = -b_1$ .
- $c_2 = I_b - J_c = I_b + J_b - I_c + I_c - J_c = -a_1 + b_1 + c_1$ .

The Fig. 18 shows the class of relation  $R([-5, 3], [2, 10], [-15, -5])$  and its inverse. The inverse is calculated as the union of the inverse of each instance of the class of relations.

- For all couple of relations  $R_1(a_1, b_1, c_1)$  and  $R_2(a_2, b_2, c_2)$  in  $S$ , we have the following property:

$$(R_1 \wedge R_2)^{-1} = R_2^{-1} \wedge R_1^{-1}$$

**Proof**

$$\begin{aligned} (R_2^{-1} \wedge R_1^{-1}) \wedge (R_1 \wedge R_2) &= (R_2(-a_2, -b_2, -a_2 + b_2 + c_2) \wedge R_1(-a_1, -b_1, -a_1 + b_1 + c_1)) \wedge R'(a_1 + a_2, b_1 + b_2, c_1 - b_2) \\ &= R''(-a_1 - a_2, -b_1 - b_2, -a_2 + b_2 + c_2 + b_1) \wedge R'(a_1 + a_2, b_1 + b_2, c_1 - b_2) = R_e(0, 0, c_2 - a_2) \\ (R_1 \wedge R_2) \wedge (R_2^{-1} \wedge R_1^{-1}) &= R'(a_1 + a_2, b_1 + b_2, c_1 - b_2) \wedge (R_2(-a_2, -b_2, -a_2 + b_2 + c_2) \wedge R_1(-a_1, -b_1, -a_1 + b_1 + c_1)) \\ &= R'(a_1 + a_2, b_1 + b_2, c_1 - b_2) \wedge R''(-a_1 - a_2, -b_1 - b_2, -a_2 + b_2 + c_2 + b_1) = R_e(0, 0, b_1 + c_1) \end{aligned}$$

- For each relation  $R(a,b,c)$  in  $S$ , we have the following property:  $(R^{-1})^{-1} = R$

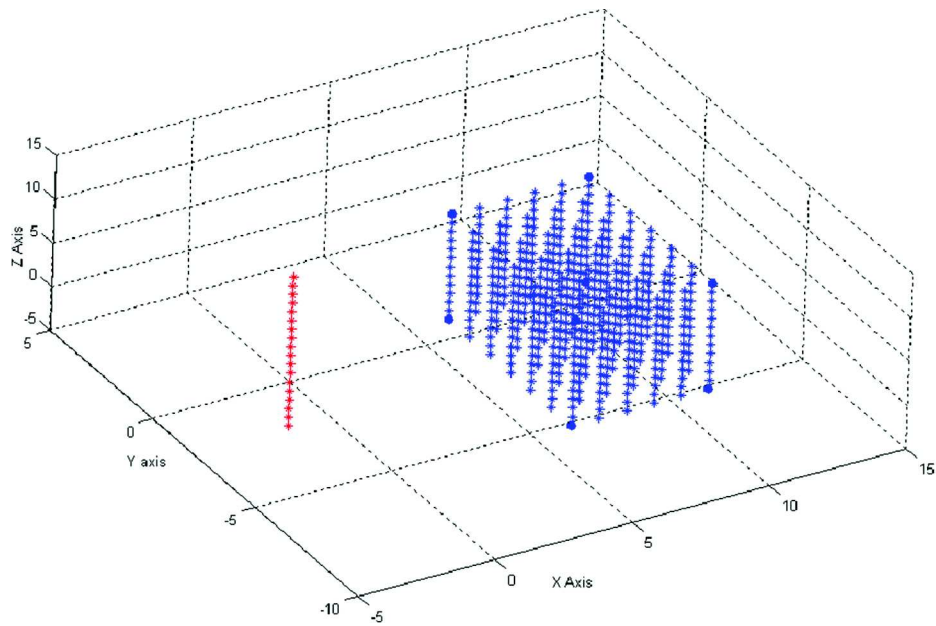


Fig. 15  $R(A,B,C)=R([5\ 10],[-1-7],[0\ 12])$  and  $R(0,0,C-A)$

**Proof**

$$\begin{aligned} \left(R(a,b,c)^{-1}\right)^{-1} &= R'(-a,-b,-a+b+c)^{-1} \\ &= R''(-(-a),-(-b),-(-a)+(-b)+(-a+b+c)) = R(a,b,c) \end{aligned}$$

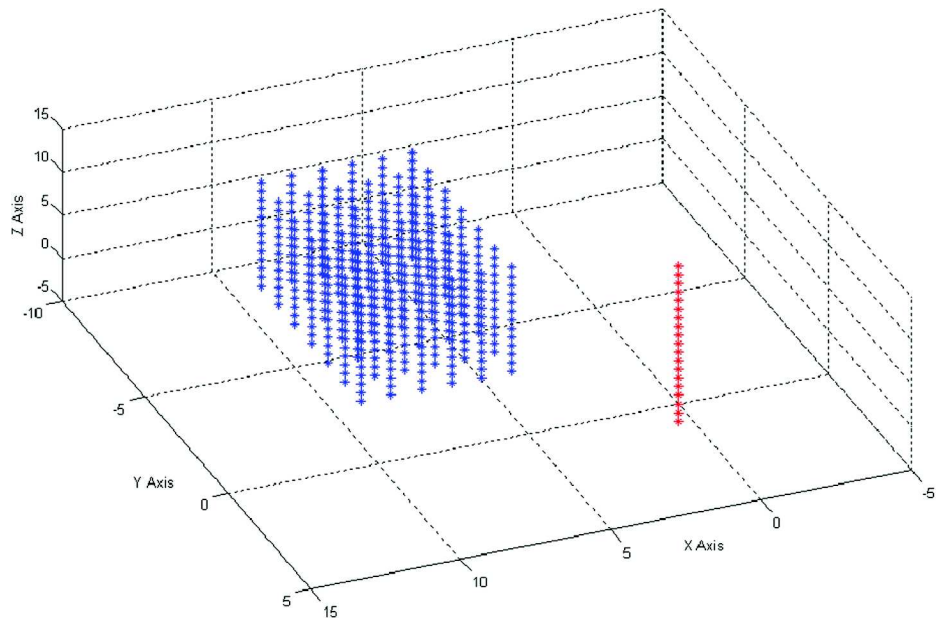


Fig. 16  $R(A,B,C)=R([5\ 10],[-1-7],[0\ 12])$  and  $R(\{0\},\{0\},B+C)$

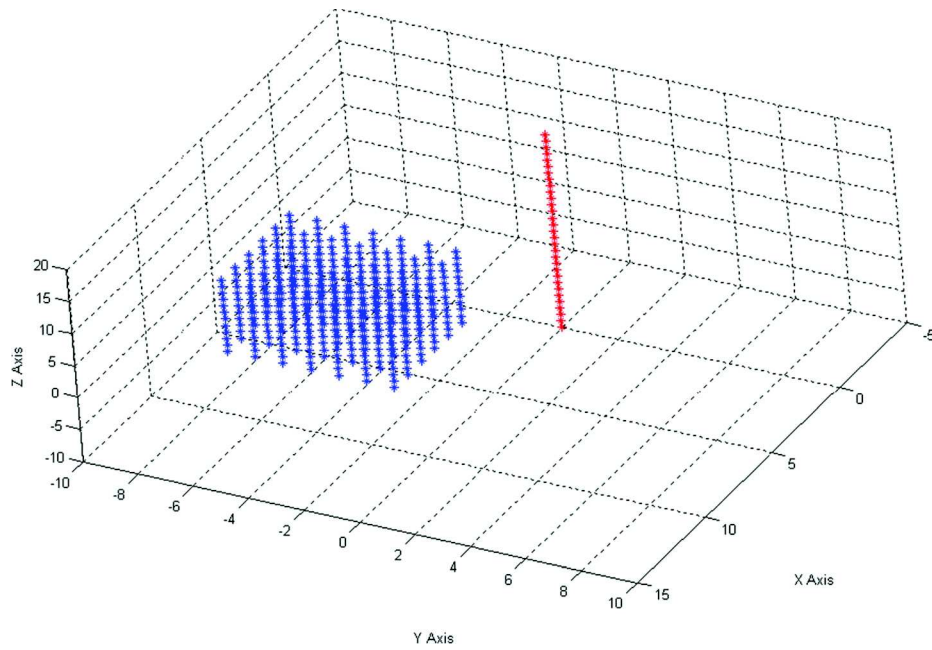


Fig. 17  $R(A,B,C)=R([5\ 10],[ -1\ -7],[0\ 12])$  and  $R(\{0\},\{0\},\mathbb{R})$

- For all relations  $R_1, R_2, R_3$  in  $S$ , we have the following property:  $R_1 = R_2 \wedge R_3 \Leftrightarrow R_2 = R_1 \wedge (R_3)^{-1} \Leftrightarrow R_3 = (R_2)^{-1} \wedge R_1$

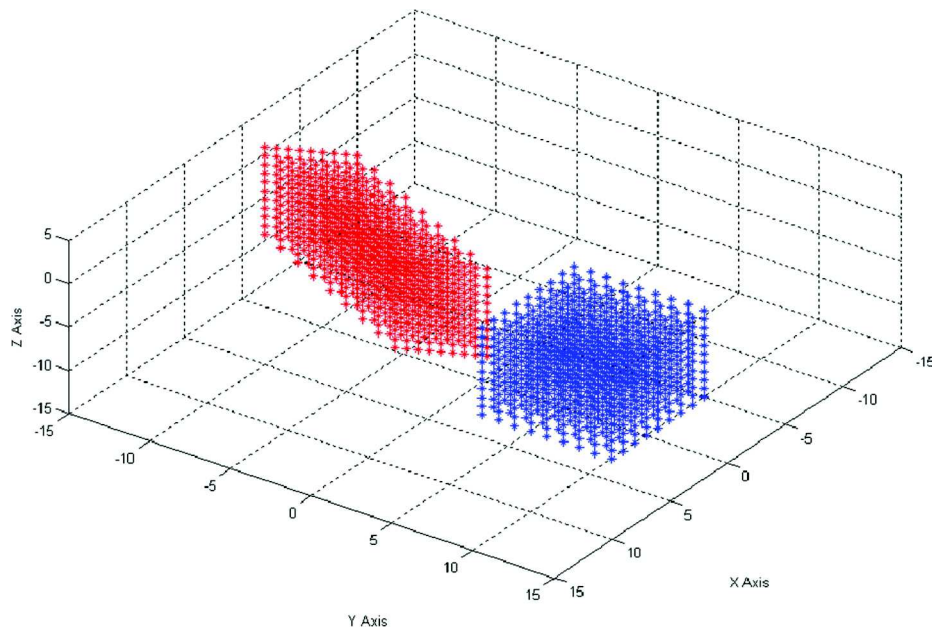


Fig. 18  $R([-5\ 3], [2\ 10], [-15\ -5])$  in blue and its inverse in red

**Proof**

$$R_1 = R_2 \wedge R_3 \Leftrightarrow R_1 \wedge R_3^{-1} = R_2 \wedge R_3 \wedge R_3^{-1} \Leftrightarrow R_1 \wedge R_3^{-1} = R_2 \wedge R_e = R_2$$

By the same way, we can verify the second property.

Using the parameters of the relations, we can verify the property as follows:

$$\begin{aligned} R_1(a_1, b_1, c_1) &= R_2(a_2, b_2, c_2) \wedge R_3(a_3, b_3, c_3) \Leftrightarrow R_1(a_1, b_1, c_1) \\ &= R_2(a_2 + a_3, b_2 + b_3, c_2 - b_3) \end{aligned}$$

In this case, we have the following equalities that hold:

- $a_1 = a_2 + a_3$
- $b_1 = b_2 + b_3$ , and
- $c_1 = c_2 - b_3$

From these equalities, we can derive the following:

- $a_2 = a_1 - a_3$
- $b_2 = b_1 - b_3$
- $c_2 = c_1 + b_3$

$$\begin{aligned} \Leftrightarrow R_2(a_2, b_2, c_2) &= R_1(a_1, b_1, c_1) \wedge R_3'(-a_3, -b_3, -a_3 + b_3 + c_3) \\ &= R_1(a_1, b_1, c_1) \wedge R_3^{-1}(a_3, b_3, c_3) \end{aligned}$$

By the same way, we can write the following:

- $a_3 = a_1 - a_2$
- $b_3 = b_1 - b_2$
- $b_3 = c_2 - c_1 \Leftrightarrow J_b - K_b = J_b - I_c - (K_b - I_c)$

After adding and subtracting variables, we will find that  $c_3 = -a_2 + b_2 + c_2 - b_1$ , which verify the property:

$$R_3(a_3, b_3, c_3) = R_2^{-1}(a_2, b_2, c_2) \wedge R_1(a_1, b_1, c_1)$$

- For all relations  $R_1, R_2, R_3$  in  $S$ , we have the following property:  $(R_1^{\sim} \vee R_2)^{\sim} \vee (R_1^{\sim} \vee R_2^{\sim})^{\sim} = R_1$

**Proof**

$$\begin{aligned} (R_1^{\sim} \vee R_2)^{\sim} \vee (R_1^{\sim} \vee R_2^{\sim})^{\sim} &= (R_1^{\sim} \cup R_2)^{\sim} \cup (R_1^{\sim} \cup R_2^{\sim})^{\sim} = ((R_1^{\sim})^{\sim} \cap R_2^{\sim}) \cup ((R_1^{\sim})^{\sim} \cap (R_2^{\sim})^{\sim}) \\ &= (R_1 \cap R_2^{\sim}) \cup (R_1 \cap R_2) = R_1 \cap (R_1 \cup R_2^{\sim}) = R_1 \end{aligned}$$

- For all relations  $R_1$  and  $R_2$  in S, we have the following property:  $(R_1 \vee R_2)^{-1} = R_1^{-1} \vee R_2^{-1}$

**Proof** For all couple of relations  $R_1(a_1, b_1, c_1)$  and  $R_2(a_2, b_2, c_2)$  in S, we have the following property:

$$\begin{aligned} (R_1 \vee R_2)^{-1} &= R'(a_1 + a_2, b_1 + b_2, c_1 + c_2)^{-1} = R(-a_1 - a_2, -b_1 - b_2, -a_1 - a_2 + b_1 + b_2 + c_1 + c_2) \\ (R_1^{-1} \vee R_2^{-1}) &= R_1(-a_1, -b_1, -a_1 + b_1 + c_1) \vee R_2(-a_2, -b_2, -a_2 + b_2 + c_2) \\ &= R(-a_1 - a_2, -b_1 - b_2, -a_1 + b_1 + c_1 - a_2 + b_2 + c_2) \\ &= R(-a_1 - a_2, -b_1 - b_2, -a_1 - a_2 + b_1 + b_2 + c_1 + c_2) \end{aligned}$$

- For all relations  $R_1, R_2, R_3$  in S, we have the following property:  $(R_1 \vee R_2) \wedge R_3 = (R_1 \wedge R_3) \vee (R_2 \wedge R_3)$

**Proof** Let  $I R(a, b, c) J$  with  $I = [I_b I_e]$  and  $J = [J_b J_e]$ . Suppose that  $R$  is an instance of  $(R_1(A_1, B_1, C_1) \vee R_2(A_2, B_2, C_2)) \wedge R_3(A_3, B_3, C_3)$ . We can write:

$$\begin{aligned} I(R_1(A_1, B_1, C_1) \vee R_2(A_2, B_2, C_2)) \wedge R_3(A_3, B_3, C_3) J & \\ \Rightarrow I(R_1(A_1, B_1, C_1) \cup R_2(A_2, B_2, C_2)) \wedge R_3(A_3, B_3, C_3) J &\Rightarrow \exists K \\ = [K_b K_e] / I (R_1(A_1, B_1, C_1) \cup R_2(A_2, B_2, C_2)) K \&\& K R_3(A_3, B_3, C_3) J \\ \Rightarrow I R_1(A_1, B_1, C_1) K \&\& K R_3(A_3, B_3, C_3) J \text{ or } I R_2(A_2, B_2, C_2) K \&\& K R_3(A_3, B_3, C_3) J & \\ \Rightarrow I (R_1(A_1, B_1, C_1) \wedge R_3(A_3, B_3, C_3)) J \text{ or } I (R_2(A_2, B_2, C_2) \wedge R_3(A_3, B_3, C_3)) J & \\ \Rightarrow I (R_1(A_1, B_1, C_1) \wedge R_3(A_3, B_3, C_3)) J \vee I (R_2(A_2, B_2, C_2) \wedge R_3(A_3, B_3, C_3)) J & \\ \Rightarrow (R_1 \vee R_2) \wedge R_3 \sqsubseteq (R_1 \wedge R_3) \vee (R_2 \wedge R_3) & \end{aligned}$$

Reciprocally, if  $I (R_1(A_1, B_1, C_1) \wedge R_3(A_3, B_3, C_3)) \vee (R_2(A_2, B_2, C_2) \wedge R_3(A_3, B_3, C_3)) J \Rightarrow I R_1(A_1, B_1, C_1) \wedge R_3(A_3, B_3, C_3) J$  or  $I R_2(A_2, B_2, C_2) \wedge R_3(A_3, B_3, C_3) J \Rightarrow \exists K = [K_b K_e] / I R_1(A_1, B_1, C_1) K \&\& K R_3(A_3, B_3, C_3) J$  or  $I R_2(A_2, B_2, C_2) K \&\& K R_3(A_3, B_3, C_3) J \Rightarrow (I R_1(A_1, B_1, C_1) K \text{ or } I R_2(A_2, B_2, C_2) K) \&\& K R_3(A_3, B_3, C_3) J \Rightarrow I (R_1(A_1, B_1, C_1) \vee R_2(A_2, B_2, C_2)) \wedge R_3(A_3, B_3, C_3) J \Rightarrow (R_1 \wedge R_3) \vee (R_2 \wedge R_3) \sqsubseteq (R_1 \vee R_2) \wedge R_3$

## 7 Operations on TRM

We have already mentioned that in the domain of multimedia content analysis, the segmentation systems provide non-convex intervals where a specific event occurs. For that reason, we have proposed a novel representation of the temporal relations between non-convex intervals that is the Temporal Relation Matrix (TRM) in addition to a novel method of analysis of such Matrix. Hence, it is important to define some of the already presented operations such as inverse, composition, disjunction but this time between TRMs rather than between two temporal relations.

### 7.1 Inverse of TRM

If we consider now occurrences of a temporal relation  $R'$  that can be observed between  $s_{2j}$  and  $s_{1i}$  ( $s_{2j} R^{-1}(DE', DB', Lap') s_{1i}$ ), we can calculate its parameters by using those of the relation  $R$  that exist between  $s_{1i}$  and  $s_{2j}$ , we can establish that

$$DE' = -DE; DB' = -DB; Lap' = -DE + DB + Lap$$

So the TRM associated to the relations that can be observed between two segmentation  $S_2$  and  $S_1$  ( $TRM(S_2, S_1)$ ) can be calculated using the  $TRM(S_1, S_2)$  by considering:

$$TRM(S_2, S_1)[i][j][k] = TRM(S_1, S_2)[-i][-j][-i + j + k].$$

## 7.2 Disjunction of two TRM

Let  $TRM_1, TRM_2$  two matrix that represent all possible relations that can present between all the point of couple of segmentations (two non-convex intervals). The disjunction of these two TRMs is the addition of the two matrices.

The value of occurrences of the disjunction of two relations is equal to the sum of the occurrences of each one.

$$TRM_1(S_1, S_2) \vee TRM_2(S_3, S_4) = TRM_1 + TRM_2$$

## 7.3 Composition of two TRM

The composition of two relations is more complex than the disjunction and the inverse operation since there are more additional constraints on the composition process. The composition of two relations should have a common interval. In other word, if we have  $IR_1J$  and  $KR_2L$ , the composition of these two relations cannot be performed unless J and K are the same intervals in the intermediate temporal segmentation. Let us consider  $S_1 = \{[s_{1ib} s_{1ie}]\}_{i \in [1 M]}$ ,  $S_2 = \{[s_{2jb} s_{2je}]\}_{j \in [1 N]}$ ,  $S_3 = \{[s_{3kb} s_{3ke}]\}_{k \in [1 P]}$  three segmentations (non-convex intervals).

Suppose  $TRM_1(S_1, S_2)$ ,  $TRM_2(S_2, S_3)$  and  $TRM_3(S_1, S_3)$  be the TRMs calculated between these non-convex intervals. We will note  $TRM = TRM_1 \cdot TRM_2$  as the composed TRM.

If we have  $S_{1i}R S_{3k}$  a relation resulted from a composition operation, it means there exists a sub-interval  $S_{2j}$  such that  $S_{1i}R S_{3k} = (S_{1i}R_1S_{2j}) \cdot (S_{2j}R_2S_{3k})$ .

Since the TRM represents the temporal relations computed between each couple  $(S_{1i}, S_{3k})$ , a couple of specific sub-intervals  $(S_{1i}, S_{3k})$  will occur N times in the composed TRM. This is because these two sub-intervals can be related through each sub-interval  $S_{2j}$  in  $S_2$ .

Therefore, it can be verified that  $TRM_1 \cdot TRM_2 = N * TRM_3$  where N = number of sub-intervals in the second non-convex interval  $S_2$ .

## 8 Application of operations on Allen's predefined relations

Through this section, we will show you as an example the application of the different operations on Allen's interval relations. We have included into the table the cases where one or two of the intervals may be points (duration=0). As shown in Tables 1, 2, and 3, the relations are transformed in the space of our representation. For example, the before relation in the Allen's interval relations is represented by the following three constraints:  $Lap > 0$ ,  $DB < -Lap$ , and  $DE > Lap$ . This relation can be written:  $before([Lap + \infty [, ] - \infty - Lap[, ]0 + \infty [)$ . However, if we take into account the fact that one of the intervals or the two intervals may be points (duration = 0), the relation can be written  $before([0 + \infty [, ] - \infty 0[, ]0 + \infty [)$ .

The following table (Table 4) shows you the different constraints of the Allen's relations.

If we take as an example the composition of the two relations *finish* and *during*. The resulted relation is the *during* relation. In other words,  $(IfJ) (JdK) = (IdK)$ . This result can be derived from our representation as follows:

$$\begin{aligned}
& f(\{0\}, ]0 + \infty[, ]-\infty 0[) \wedge d(]0 + \infty[, ]0 + \infty[, ]-\infty 0[) \\
&= R_1(\{0\} + ]0 + \infty[, ]0 + \infty[ + ]0 + \infty[, ]-\infty 0[-]0 + \infty[) \cap R_2(\{0\} + ]0 + \infty[, ]0 \\
&+ \infty[ + ]0 + \infty[, \{0\} + ]-\infty 0[) \\
&= R_1(]0 + \infty[, ]0 + \infty[, ]-\infty 0[) \cap R_2(]0 + \infty[, ]0 + \infty[, ]-\infty 0[) \\
&= R(]0 + \infty[, ]0 + \infty[, ]-\infty 0[)
\end{aligned}$$

By the same way, the composition of the relations *si* (*start inverse*) and *before* is computed as follows:

$$\begin{aligned}
& s_i(]-\infty 0[, \{0\}]-\infty 0[) \wedge b(]0 + \infty[ ]-\infty 0[ ]0 + \infty[) \\
&= R_1(]-\infty 0[ + ]0 + \infty[, \{0\} + ]-\infty 0[ ]-\infty 0[-] -\infty 0[) \cap R_2(]-\infty 0[ + ]0 \\
&+ \infty[, \{0\} + ]-\infty 0[ ]-\infty 0[ + ]0 + \infty[) \\
&= R_1(]-\infty 0[ + \infty[ ]-\infty 0[ ]-\infty 0[ + \infty[) \cap R_2(]-\infty 0[ + \infty[ ]-\infty 0[ ]-\infty 0[ + \infty[) \\
&= R(]-\infty 0[ + \infty[ ]-\infty 0[ ]-\infty 0[ + \infty[)
\end{aligned}$$

The constraints of the resulted relation R is the union of the constraints of several relations among them the relations *before* (<), *meet* (m), *overlap* (o), *finish inverse* (fi) and *during inverse* (di). The remaining relations are not associated to any relation in the space of Allen's relations. Table 5 shows the different possible relations issued from the composition of two basic relations. They are split over two tables for simplicity.

To validate the inverse operation, let us take the relation *finish* represented by *finish* ( $\{0\}, ]0 + \infty[, ]-\infty 0[$ ). As we have already presented, the inverse of a relation  $R(A, B, C)$  is  $R^{-1}(-A, -B, -A + B + C)$ . Hence, the inverse of the finish relation is  $R(-\{0\}, -(]0 + \infty[), -(\{0\} + ]0 + \infty[) + (]-\infty 0[)) = R(\{0\}, ]-\infty 0[, ]-\infty 0[ + \infty[)$ . The resulted relation is the union of the following relations:  $R_1(\{0\}, ]-\infty 0[, ]-\infty 0[ + \infty[)$ ,  $R_2(\{0\}, ]-\infty 0[, \{0\})$ , and  $R_3(\{0\}, ]-\infty 0[, ]0 + \infty[)$ .  $R_1$  is the relation *finish inverse* (fi) while  $R_2$  and  $R_3$  are not associated to any relation in the Allen's space.

## 9 Application of operations on clustered relations

We have already presented two ways to analyze a TRM. The first by decomposing the 3D space represented by the TRM into predefined zones each associated to a semantic relation such as the Allen's relations. The second way generically proposes to base on the distribution of votes in order to derive the relations named classes of temporal relations. A clustering algorithm such as k-means, hierarchical or any other clustering algorithm can be used to produce such classes of relations. When classes of relations are ready, each is represented by the number of votes included in the cluster in addition to other parameters if necessary.

When a clustering algorithm is used to highlight classes of relations in the TRM, one of two methods can be used in order to represent a class of relations. In the first method, each class is represented by the constraints of the 3D zone that includes all the points of this class in addition to the number of votes contained in the zone. Using this method, the previously defined operations can be used by the same way as in the case of Allen's relations (refer to Section 5). While the second method represents each cluster by its mean (or median) and its covariance matrix. In the latter case, a class of relations is noted  $R(m, \Sigma, nb)$  where m is the mean,  $\Sigma$  is the covariance matrix and nb is the number of occurrences of relation instances in the cluster.

**Table 4** Constraints associated to Allen's and villain and Kautz temporal relations

Relation	DE		DB		Lap	
<	]0	+∞[	]-∞	0 [	]0	+∞[
m	]0	+∞[	]-∞	0[	{0}	
o	]0	+∞[	]-∞	0[	]-∞	0[
s	]0	+∞[	{0}		]-∞	0[
f	{0}		]0	+∞[	]-∞	0[
=	{0}		{0}		]-∞	0[
d	]0	+∞[	]0	+∞[	]-∞	0[
>	]-∞	0[	]0	+∞[	]DE - DB	+∞[
mi	]-∞	0[	]0	+∞[	{DE - DB}	
oi	]-∞	0[	]0	+∞[	]-∞	DE - DB[
si	]-∞	0[	{0}		]-∞	0[
fi	{0}		]-∞	0[	]-∞	0[
di	]-∞	0[	]-∞	0[	]-∞	0[

### 9.1 Inverse of a class of relations

To compute the inverse of a class of relations  $R$ , we compute the inverse of each instance  $R(a,b,c)$  in  $R$  as we have already presented ( $R^{-1}(-a, -b, -a + b + c)$ ). Then the mean and the covariance matrix are calculated for the new class of relations while the number of occurrences remains the same. The parameters of the inverse relation can be also derived from the parameters of the relation  $R$  directly as follows:

Let  $R(m, \Sigma, nb)$  be a class of relations represented in addition to the number of occurrences by the following parameters:

$m = (m_x, m_y, m_z)$  is the mean value.

$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{xy} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{xz} & \Sigma_{yz} & \Sigma_{zz} \end{bmatrix}$  is the covariance matrix.

The inverse relation is  $R^{-1}(m^{-1}, \Sigma^{-1}, nb)$  where:  $m^{-1} = (-m_x, -m_y, -m_x + m_y + m_z)$

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xx} - \Sigma_{xy} - \Sigma_{xz} \\ & \Sigma_{yy} & \Sigma_{xy} - \Sigma_{yy} - \Sigma_{xz} \\ & & \Sigma_{xx} + \Sigma_{yy} + \Sigma_{zz} - 2\Sigma_{xy} - 2\Sigma_{xz} + 2\Sigma_{yz} \end{bmatrix}$$

### 9.2 Disjunction of two classes of relations

The disjunction of two classes of relations is approximated as the union of the means and the co-variance matrices of two classes of relations produced by clustering algorithm.

$$R_1(m_1, \Sigma_1, nb_1) \vee R_2(m_2, \Sigma_2, nb_2) = R(\{m_1, m_2\}, \{\Sigma_1, \Sigma_2\}, \{nb_1, nb_2\})$$

### 9.3 Composition of two classes of relations

The composition of two classes of relations is more complicated than the two operations presented above. The complication comes from the fact that the composition of two relations  $I R_1 J$  and  $K R_2 L$  has no sense unless  $J$  and  $K$  are the same interval. When we construct the TRMs and then we cluster data into classes of relations into each TRM, we do not keep track of the



**Table 5** Composition table of Allen's relations, "all" means the whole set of relations

-	<	m	o	s	f	=		
<	<	<	<	<	{<,m,o,s,d}	<		
m	<	<	<	m	{o,s,d}	m		
o	<	<	{<,m,o}	o	{o,s,d}	o		
s	<	<	{<,m,o}	s	d	s		
f	<	m	{o,s,d}	d	f	f		
=	<	m	o	s	f	=		
d	<	<	{<,m,o,s,d}	d	d	d		
>	all	{f,d,mi,oi}	{f,d,mi,oi}	{f,d,mi,oi}	>	>		
mi	{<,m,o,fi,di}	{s,=,si}	{f,d,oi}	{d,f,oi}	mi	mi		
oi	{<,m,o,fi,di}	{o,fi,di}	{o,s,f,=,d,oi,si,fi,di}	{d,f,oi}	oi	oi		
si	{<,m,o,fi,di}	{o,fi,di}	{o,fi,di}	{s,=,si}	oi	si		
fi	<	m	o	o	{f,=,fi}	fi		
di	{<,m,o,fi,di}	{o,fi,di}	{o,fi,di}	{o,fi,di}	{oi,si,di}	di		
-	d	>	mi	oi	si	fi		di
<	{<,m,o,s,d}	all	{<,m,o,s,d}	{<,m,o,s,d}	{<}	{<}		{<}
m	{o,s,d}	{di,si,oi,mi, >}	{=,f,fi}	{o,s,d}	{m}	{<}		{<}
o	{o,s,d}	{di,si,oi,mi, >}	{oi,si,di}	{o,s,f,=,d,oi,si,fi,di}	{o,fi,di}	{<,m,o}		{<,m,o,fi,di}
s	{d}	{>}	{mi}	{f,d,oi}	{s,=,si}	{<,m,o}		{<,m,o,fi,di}
f	{d}	{>}	{>}	{>,mi,oi}	{>,mi,oi}	{f,=,fi}		{>,mi,oi,si,di}
=	{d}	{>}	{mi}	{oi}	{si}	fi		{di}
d	{d}	{>}	{>}	{f,d,mi,oi,>}	{f,d,mi,oi,>}	{<,m,o,s,d}		all
>	{f,d,mi,oi}	{>}	{>}	{>}	{>}	{>}		{>}
mi	{f,d,oi}	{>}	{>}	{>}	{>}	{mi}		{>}
oi	{f,d,oi}	{>}	{>}	{>,mi,oi}	{>,mi,oi}	{oi,si,di}		{>,mi,oi,si,di}
si	{f,d,oi}	{>}	{mi}	{oi}	{si}	{di}		{di}
fi	{o,s,d}	{di,si,oi,mi, >}	{oi,si,di}	{oi,si,di}	{di}	{fi}		{di}
di	{o,s,f,=,d,oi,si,fi,di}	{di,si,oi,mi, >}	{oi,si,di}	{oi,si,di}	{di}	{di}		{di}

intervals that have voted for each relation in the TRM. Thus, the composition of two instances of relations one from each cluster may have a common interval and may not have. However, the main idea of our representation is to highlight the relations that may occur frequently or rarely compared to other relations and hence the number of occurrences has the main importance for us. To overcome the problem of composition, we proceed to an approximation of the composition of two classes of relations.

Let us consider the following two classes of relations that we want to compute their composition.

$$C_1 = \left\{ \left( p_{ix}, p_{iy}, p_{iz} \right), i = 1 \dots M \right\}$$

$$C_2 = \left\{ \left( q_{jx}, q_{jy}, q_{jz} \right), j = 1 \dots N \right\}$$

$$C = C_1 \wedge C_2 = \left\{ \left( r_{kx}, r_{ky}, r_{kz} \right), k = 1 \dots P \right\}$$

Each of the above classes is represented by the mean of the relation instances in the class, the covariance matrix and the number of instances in the class. We will note them as follows:

$$C_1(m_{c1}, \Sigma^{c1}, M) / m_{c1} = (m_{1x}, m_{1y}, m_{1z}), \quad C_2(m_{c2}, \Sigma^{c2}, N) / m_{c2} = (m_{2x}, m_{2y}, m_{2z}), \quad C(m_c, \Sigma_c, P) / m_c = (m_x, m_y, m_z)$$

$$\Sigma^{c1} = \begin{bmatrix} \Sigma_{xx}^{c1} & \Sigma_{xy}^{c1} & \Sigma_{xz}^{c1} \\ \Sigma_{xy}^{c1} & \Sigma_{yy}^{c1} & \Sigma_{yz}^{c1} \\ \Sigma_{xz}^{c1} & \Sigma_{yz}^{c1} & \Sigma_{zz}^{c1} \end{bmatrix}, \quad \Sigma^{c2} = \begin{bmatrix} \Sigma_{xx}^{c2} & \Sigma_{xy}^{c2} & \Sigma_{xz}^{c2} \\ \Sigma_{xy}^{c2} & \Sigma_{yy}^{c2} & \Sigma_{yz}^{c2} \\ \Sigma_{xz}^{c2} & \Sigma_{yz}^{c2} & \Sigma_{zz}^{c2} \end{bmatrix}, \quad \Sigma^c = \begin{bmatrix} \Sigma_{xx}^c & \Sigma_{xy}^c & \Sigma_{xz}^c \\ \Sigma_{xy}^c & \Sigma_{yy}^c & \Sigma_{yz}^c \\ \Sigma_{xz}^c & \Sigma_{yz}^c & \Sigma_{zz}^c \end{bmatrix}$$

In the ideal case, the composition of two classes is computed by performing the operation between two points each one comes from a class providing that these two points be computed with a common interval. Otherwise, the relation will not be defined.

To ensure that the two classes will have points calculated between common intervals, we compose only classes that results from TRMs with a common non-convex interval (same segmentation). In other words, suppose  $C_1$  ( $C_2$  resp.) is a class resulted from the clustering of data in  $TRM_1$  ( $TRM_2$  resp.) with  $TRM_1$  ( $TRM_2$  resp.) computed between the two non-convex intervals  $S_1$  and  $S_2$  ( $S_3$  and  $S_4$  resp.). The composition of  $C_1$  and  $C_2$  is defined only in the case where  $S_2$  and  $S_3$  are the same non-convex interval. Only in this case, we are sure that some relation instances in  $C_1$  can be composed with some relation instances in  $C_2$ . However, we cannot know each relation instance in  $C_1$  can be composed with which relation instances in  $C_2$ . Thus, we proceed to an approximation of the composition. The approximation is done by composing each relation instance in  $C_1$  with each relation instance in  $C_2$  even though some of them will not have a common interval. The parameters of the resulted class can be derived from the parameters of  $C_1$  and  $C_2$  as described below.

$$F = (f_1, f_2) : ((m_x, \Sigma_x), (m_y, \Sigma_y)) \rightarrow (m_z, \Sigma_z) / m_z = f_1(m_x, m_y), \Sigma_z = f_2(\Sigma_x, \Sigma_y)$$

$$m_z \approx (m_{x1} + m_{y1}, m_{x2} + m_{y2}, m_{x3} - m_{y2} = m_{x1} + m_{y3})$$

$$\Sigma_z \approx \begin{bmatrix} \Sigma_{xx}^{c1} + \Sigma_{xx}^{c2} & \Sigma_{xy}^{c1} + \Sigma_{xy}^{c2} & \Sigma_{xz}^{c1} - \Sigma_{xy}^{c2} \\ \Sigma_{xy}^{c1} + \Sigma_{xy}^{c2} & \Sigma_{yy}^{c1} + \Sigma_{yy}^{c2} & \Sigma_{yz}^{c1} - \Sigma_{yy}^{c2} \\ \Sigma_{xz}^{c1} - \Sigma_{xy}^{c2} & \Sigma_{yz}^{c1} - \Sigma_{yy}^{c2} & \Sigma_{zz}^{c1} + \Sigma_{yy}^{c2} \end{bmatrix}$$

In order to test the approximation that we have done, we have completed it by an experimentation using simulated data. Given two classes  $C_1$  and  $C_2$  generated randomly using a Gaussian distribution, the parameters of the composed class  $C = C_1 \wedge C_2$  can be computed in two ways as mentioned before. The first way is to do a point-by-point composition between the two classes and then compute the parameters of the obtained result. In the second way, the parameters of  $C$

**Table 6** Comparison between approximated and non-approximated parameters

Covariance Matrix of $C_1$			Covariance Matrix of $C_2$		
30,514	-29,421	30,241	24,756	-24,236	24,292
-29,421	29,563	-29,381	-24,236	24,558	-24,337
30,241	-29,381	30,280	24,292	-24,337	24,404
Approximated Covariance Matrix of C			Covariance Matrix of C using Point-by-Point Composition		
55,269	-53,657	54,477	55,018	-53,414	54,228
-53,657	54,121	-53,939	-53,414	53,877	-53,696
54,477	-53,939	54,838	54,228	-53,626	54,589
Mean of $C_1$			Mean of $C_2$		
3.3161	-10.852	-25.839	12.08	-9.3222	-15.498
Approximated Mean of C			Mean of C using Point-by-Point Composition		
15.396	-20.174	-16.516	15.396	-20.174	-16.516

are derived from the parameters of the two classes by approximation as described above. Table 6 lists the obtained results.

As we can notice, the approximated values of the parameters of C are very close to the ones calculated after a point-by-point composition of  $C_1$  and  $C_2$ .

Figure 19 shows two classes to be composed and Fig. 20 shows the composed class using a point-by-point composition.

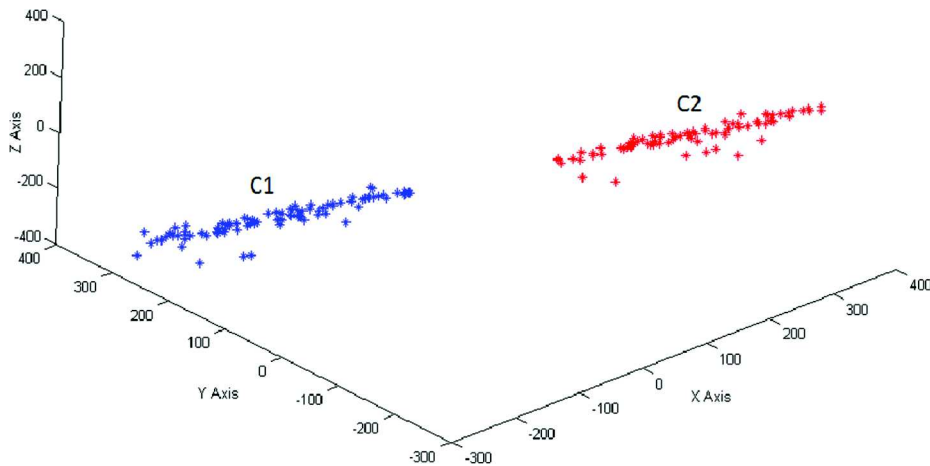
Based on the theorem concerning the relations algebras [58], we can deduce that:

$(S, \cap, \cup, C_R, \emptyset, \mathcal{R}, \wedge, ^{-1}, R_e)$  is an algebra of temporal relations, where:

$S = P(\mathcal{R})$  = the set of partitions of  $\mathbb{R}^3$ ,  $C_R$  the complement of a relation R,  $R(\emptyset, \emptyset, \emptyset)$  the empty relation,  $\mathcal{R} = (\mathbb{R}, \mathbb{R}, \mathbb{R})$  the total relation,  $^{-1}$  the inverse operator, and  $R_e(\{0\}, \{0\}, \mathbb{R})$  the identity relation.

## 10 Video analysis using the defined temporal reation algebra

The analysis of the temporal relations that exist between different segmentations has been used in several already-published works for audio-visual document structuring, event detection and audio-visual document classification [27–29].



**Fig. 19** The two classes to be composed

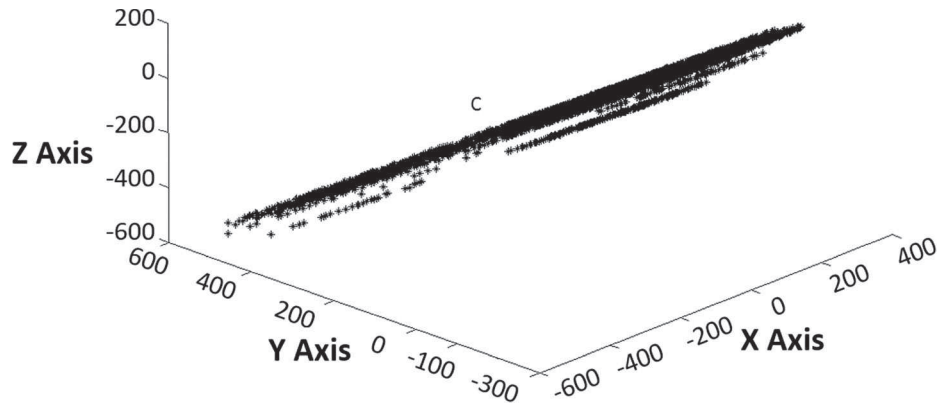


Fig. 20 The composition of  $C_1$  and  $C_2$

In [27], the analysis of French TV-game video is done in order to highlight the structure of the video. The analysed video document is a 31 min' video. From an audio-point of view, the video is a competition between two teams, each with two players (speaker #2, speaker#3) and (speaker #4, speaker #5). Two animators are animating the TV game, speaker #1 the principal animator while speaker #6 and speaker #7 are secondary ones (presenting the audiences that participate to the game and the set of gifts to be won). One audience is also appearing in the program (speaker #8). The video is composed of several game phases each is a sequence of interaction between the two speakers of the same team. When a team wins a game phase, the audiences start applauding. Eight elementary segmentations are derived from the speaker segmentation, one per speaker. A TRM is computed between each couple of elementary segmentations, the optimal number of clusters is determined in each TRM and the total number of votes in each cluster is calculated. A parameter is introduced in the process of TRM calculation to consider only the temporal relations between the two segments of two segmentations having a gap less than 10 s. In other words, if the two segments are too far from each other, the temporal relation between them will be less relevant and will be discarded. Table 7 shows the distribution of votes between clusters in each TRMs. Two clusters are identified,  $C_1$  represents the case when the first speaker of the first segmentation is speaking to the second one while the second cluster  $C_2$  represents the inverse case. In this table,  $TRM_{S(1,2)}$  stands for the TRM computed between speaker #1 and speaker #2.

Table 7 Distribution of votes between clusters in TRMs ([12])

TRM	C1	C2	TRM	C1	C2	TRM	C1	C2
$TRM_{S(1,2)}$	65	60	$TRM_{S(1,3)}$	49	49	$TRM_{S(1,4)}$	84	71
$TRM_{S(1,5)}$	106	97	$TRM_{S(1,6)}$	6	5	$TRM_{S(1,7)}$	89	79
$TRM_{S(1,8)}$	3	5	$TRM_{S(2,3)}$	123	124	$TRM_{S(2,4)}$	4	7
$TRM_{S(2,5)}$	6	6	$TRM_{S(2,6)}$	0	0	$TRM_{S(2,7)}$	6	7
$TRM_{S(2,8)}$	0	0	$TRM_{S(3,4)}$	6	5	$TRM_{S(3,5)}$	10	5
$TRM_{S(3,6)}$	0	0	$TRM_{S(3,7)}$	7	4	$TRM_{S(3,8)}$	0	0
$TRM_{S(4,5)}$	245	205	$TRM_{S(4,6)}$	4	8	$TRM_{S(4,7)}$	15	19
$TRM_{S(4,8)}$	0	0	$TRM_{S(5,6)}$	0	0	$TRM_{S(5,7)}$	39	26
$TRM_{S(5,8)}$	0	0	$TRM_{S(6,7)}$	1	0	$TRM_{S(6,8)}$	0	3
$TRM_{S(7,8)}$	4	3						

**Table 8** Distribution of votes between clusters in TRMs ([12])

TRM	C1	C2	C3	TRM	C1	C2	C3	TRM	C1	C2	C3
TRM <sub>F(1,2)</sub>	25	15	0	TRM <sub>F(1,3)</sub>	10	10	0	TRM <sub>F(1,4)</sub>	19	19	5
TRM <sub>F(1,5)</sub>	19	13	2	TRM <sub>F(1,6)</sub>	9	19	0	TRM <sub>F(1,7)</sub>	2	1	8
TRM <sub>F(1,8)</sub>	2	1	0	TRM <sub>F(2,3)</sub>	38	36	53	TRM <sub>F(2,4)</sub>	11	7	0
TRM <sub>F(2,5)</sub>	3	2	0	TRM <sub>F(2,6)</sub>	6	6	0	TRM <sub>F(2,7)</sub>	2	1	0
TRM <sub>F(2,8)</sub>	0	0	0	TRM <sub>F(3,4)</sub>	1	2	0	TRM <sub>F(3,5)</sub>	1	0	0
TRM <sub>F(3,6)</sub>	4	4	0	TRM <sub>F(3,7)</sub>	1	0	0	TRM <sub>F(3,8)</sub>	0	0	0
TRM <sub>F(4,5)</sub>	52	50	76	TRM <sub>F(4,6)</sub>	14	5	0	TRM <sub>F(4,7)</sub>	4	1	0
TRM <sub>F(4,8)</sub>	4	5	0	TRM <sub>F(5,6)</sub>	22	14	0	TRM <sub>F(5,7)</sub>	1	1	0
TRM <sub>F(5,8)</sub>	4	5	0	TRM <sub>F(6,7)</sub>	2	2	0	TRM <sub>F(6,8)</sub>	1	0	0
TRM <sub>F(7,8)</sub>	2	1	0								

As we can notice, we have high interactions between speakers #2 and #3 (first team) and between speakers #4 and #5 (second team). Moreover, we can highlight that the first speaker has considerable number of interactions with almost all the other speakers due to his role as animator. The composition of the TRM<sub>S(2,3)</sub> (TRM<sub>S(4,5)</sub> resp.) with itself several times highlights the game phases of the video. Composing the obtained game phases with the applause segmentations will highlight the won game phases only.

Another type of segmentations has been considered in this work. The face detection and recognition process provides eight elementary segmentations each contains the segments where the same face appears on the screen. The TRMs between each couple of segmentations is computed as above and the optimal number of clusters is determined. The optimal number of clusters in almost all the TRM is two while it is equal to three in TRM<sub>F(1,4)</sub>, TRM<sub>F(1,5)</sub>, TRM<sub>F(4,5)</sub>, TRM<sub>F(2,3)</sub> and TRM<sub>F(1,7)</sub>. C1 is the cluster when the first face appears before the second, C2 the inverse clusters, and C3 is the cluster when the two faces appears at the same time on the screen. Table 8 provides the distribution of votes between clusters in each TRM.

**Table 9** Experimental corpus description

Document type	Subtype	Number	duration		
			Mean	Min	Max
TV news: TREC2003	ABC	49	00:34:20	00:32:50	00:35:50
	CNN	49	00:34:20	00:32:00	00:37:20
TV news: TREC2004	ABC	6	00:34:00	00:33:50	00:34:10
	CNN	6	00:34:40	00:34:10	00:35:55
TV news: TREC2005	CCTV	9	00:53:50	00:34:00	01:10:00
	CNN	7	00:54:30	00:34:00	01:10:00
	LBC	8	00:54:40	00:30:00	01:10:00
	NBC	7	00:33:00	00:28:40	00:34:00
	MSNBC	6	00:34:00	00:34:00	00:34:00
	NTDTV	4	00:34:00	00:34:00	00:34:00
TV news: Argos	France2	17	00:39:40	00:24:40	00:44:40
Soccer game sequenes		20	01:26:10	00:25:00	02:45:00
Documentary films		21	00:29:10	00:14:00	01:05:10
TV Series	Stargate	24	00:42:18	00:39:55	00:42:20
French TV games	Les'amours	5	00:31:55	00:30:30	00:36:00
Movie extracts	Matrix	4	00:31:20	00:24:00	00:38:30
Total duration		242		6d:6 h:4 m:27 s	

**Table 10** F-measure of the k-means clustering - 6 clusters

Clustering results	F-measure (%)	Miss-classified in the class	Miss-classified out the class
News	99.1%	0	3/168
Soccer	91.9%	0	3/20
TV Series	92.33%	4	0/24
Documentary	97.7%	0	1/21
Tv Games	88.9%	0	1/5
Movie extracts	0%	0	4/4

The distribution of votes in Table 8 validates the results obtained when considering the speaker segmentations. Moreover, such results can be used to associate faces to speakers.

The distribution of votes over the clusters in each TRM was used in [29] in order to compute the similarity between audiovisual documents. Several elementary segmentations derived from dominant color, motion quantity, contrast, speakers, faces, applauses, speech, music, silence are used. Each video  $v$  is represented by the set of the TRMs computed between each couple of elementary segmentations.

$$TRM^v = \{TRM_1^v, TRM_2^v, TRM_3^v, \dots, TRM_M^v\}$$

On each  $TRM_i^v$ , a clustering method is applied in order to highlight automatically how votes are distributed in the co-occurrence matrix and the number of votes (NbV) in each cluster is counted.

$$TRM_i^v = \{NbV_{i,1}^v, NbV_{i,2}^v, NbV_{i,3}^v, \dots, NbV_{i,K}^v\}$$

So, a video  $v$  is represented by a matrix  $M^v$  of numeric values in which each row contains the number of votes of the clusters highlighted in one TRM. A distance between two videos  $v_1$  and  $v_2$  is defined as a weighted distance between their two matrices  $M^{v_1}$  and  $M^{v_2}$  as follows:

$$d(v_1, v_2) = \sum_{i=1}^M \alpha_i \left[ \sum_{j=1}^K \beta_j \left| \frac{NbV_{i,j}^{v_1}}{t_{v_1}} - \frac{NbV_{i,j}^{v_2}}{t_{v_2}} \right| \right]$$

Where  $t_{v_1}$  and  $t_{v_2}$  are the time durations of  $v_1$  and  $v_2$ .

Several supervised and unsupervised classification methods have been tested using the dataset shown in Table 9.

Table 10 shows the clustering results obtained using the k-means algorithm with  $k = 6$ .

As supervised classification methods, we have defined a simple supervised method and tested a set of well-known classifiers. The proposed supervised method takes as input a set of

**Table 11** F-measure of the proposed supervised method

10% models training	F-measure (%)	Miss-classified in the class	Miss-classified out the class
News	99.1%	0	3/168
Soccer	94.8%	0	2/20
TV Series	91.32%	1	3/24
Documentary	95.45%	2	0/21
Tv Games	80%	1	1/5
Movie extracts	61.6%	5	0/4

**Table 12** Three-folds cross-validation sampling method—F-measure

Cross validation: 3- folds	F-measure (%)					
	News	Soccer	TV series	Documentary	TV games	Movie extracts
Random forest	98.23	100	86.96	100	61.54	40
C4.5	96.68	82.05	76.6	93.33	57.14	50
Classification Tree	97.9	81.08	82.35	97.67	72.73	22.22
SVM	99.40	94.74	82.35	84.45	88.89	40
CN2 Rules	96.83	88.89	84.45	95.45	33.33	66.67
KNN	99.7	94.74	92.30	90.48	88.89	75
Naïve Bayes	98.49	97.44	93.33	89.36	88.89	61.53

videos of each category. Each video is represented by a matrix of numeric values  $M$  as stated above. Then, the method computes simply the average matrix  $M$  of all the matrices of the videos in each category and takes them as the models of the categories. To classify a video  $v$ , the distance between its associated matrix  $M^v$  and the matrix of each category is computed. The label of the category having the least distance with the video matrix is assigned. Table 11 shows the results obtained by applying the proposed supervised method trained on 10% of the dataset and tested on the remaining 90% while Table 12 shows the results obtained by applying a set of well-known classification algorithms using a 3-folds cross validation sampling method.

A pioneer idea to be tested for video classification is to train a 3D CNN on the set of TRMs. We can consider to train one 3D CNN per TRM or feed the set of TRMs in one 3D CNN.

A third type of applications of the temporal relations analysis is the conversation detection process proposed in [28]. The speaker segmentations are used in order to highlight high interactions between speakers. The TRM between each couple of speaker segmentations  $speaker_i$  and  $speaker_j$  is computed. The analysis of the TRMs highlights the interactions between speakers of the form “ $speaker_i / speaker_j$ ” and “ $speaker_j / speaker_i$  (the symbol / stands for interacts). Then, the composition operator is applied several times in order to detect the patterns “ $speaker_i / speaker_j / speaker_i / speaker_j \dots$ ” or “ $speaker_i / speaker_j / speaker_i / speaker_k \dots$ ” which may correspond to a conversation between the involved speakers. The proposed method is validated on a 65-min debate video. The method is analyzed deeply in the work of bigot et al. in [6] based on the interactions between speakers.

## 11 Conclusion

In this article, we have proposed a novel temporal model based on parametric representation of temporal relations between convex segments. Then, we have proposed a new representation of temporal relations between two non-convex segments and we called it Temporal Relation Matrix (TRM). The analysis of such TRMs allowed us to highlight classes of temporal relations. Operations such as composition and disjunction are defined in the framework of temporal relation algebra that are later applied on the classes of relations in order to highlight events of higher level in the video analysis domain. The defined temporal relations algebra is applied on different audiovisual documents in order to highlight clues about the structure of the documents, to detect events or to categorize documents in predefined or non-predefined types. However, it will be very interesting to study the effectiveness of the temporal algebra in order

to analyze data coming from other domains or for other type of applications such as the constraint satisfaction problem (CSP). Moreover, and as stated in the introduction, we aim to use deep learning techniques to learn video structures by feeding the TRMs into 3D CNNs.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Allen JF (1983) Maintaining knowledge about temporal intervals. *J Commun ACM* 26(11):832–843
2. Anant B, Cho J, Lee W, Ko B-S (2015) Sports highlights generation based on acoustic events detection: a rugby case study. In: *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, USA
3. Avrithis Y, Tsapatsoulis N, Kollias S (2000) Broadcast news parsing using visual cues: a robust face detection approach. In: *IEEE International Conference on Multimedia and Expo (ICME2000)*, New York, USA
4. Balbiani P, Osmani A (1999) Représentation et Raisonnement sur les Intervalles Cycliques. In: *Journées nationales sur les modèles de raisonnement (JNMR)*, France
5. Balbiani P, Condotta J-F, Ligozat G (2003) Reasoning about cyclic Space: axiomatic and computational aspects. In: *Spatial Cognition III (SC 2002)*, Allemagne
6. Bigot B, Pinquier J, Ferrane I, Andre-Obrecht R (2012) Detecting individual role using features extracted from speaker diarization results. *Multimed Tools Appl* 60(2):347–369
7. Bonzanini A, Leonardi R, Migliorati P (2001) Exploitation of temporal dependencies of descriptors to extract semantic information. In: *International Conference on Very Low Bitrate Video Coding (VLBV2001)*, Athens, Greece
8. Buchanan C, Zellweger P (1993) Automatic temporal layout mechanisms. In: *ACM International Conference on Multimedia*, California, USA
9. Chittaro L, Montanari A (1996) Trends in temporal representation and reasoning. *Knowl Eng Rev* 11(3): 281–288
10. Chittaro L, Montanari A (2002) Temporal representation and reasoning in artificial intelligence: issues and approaches. *Ann Math Artif Intell* 28(1–4):47–106
11. Condotta J-F (2000) *Problèmes de Satisfaction de Contraintes Spatiales: Algorithmes et Complexité*. Institut de Recherche en Informatique de Toulouse, Toulouse
12. Cukierman D, Delgrande J (2004) A theory for convex interval relations including unbounded intervals. In: *International Florida Artificial Intelligence Research Society Conference*, Florida, USA
13. Dechter R, Meiri I, Pearl J (1991) Temporal constraint networks. *Artif Intell* 49(1–3):61–95
14. Dingeldein D (1994) Modeling multimedia objects with MME. In: *Eurographics Workshop on Object Oriented Graphics*, Sintra, Portugal
15. Donahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Saenko K, Darrell T (2017) Long-term recurrent convolutional networks for visual recognition and description. *IEEE Trans Pattern Anal Mach Intell* 39(4):677–691
16. Duan L-Y, Xu M, Tian Q, Xu C-S, Jin J (2005) A unified framework for semantic shot classification in sports videos. In: *IEEE Transactions on Multimedia*, Juan-les-Pins, France
17. Duda A, Keramane C (1995) Structured temporal composition of multimedia data. In: *IEEE International Workshop on Multimedia Database Management Systems*, New York, USA
18. Eickeler S, Muller S (1999) Content-based video indexing of TV broadcast news using hidden Markov models. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP1999)*
19. Freksa C (1992) Temporal reasoning based on semi-intervals. *Artif Intell* 54(1–2):199–227
20. Geng Y, Zhang G, Li W, Gu Y, Liang R-Z, Liang G, Wang J, Wu Y, Patil N, Wang J-Y (2017) A novel image tag completion method based on convolutional neural transforms. In: *International Conference on Artificial Neural Networks*, Alghero, Italy
21. M. Golumbic and R. Shamir, "Complexity and algorithms for reasoning about time: a graph-theoretic approach," *J ACM*, vol. 40, no. 5, pp. 1108–1133, November 1993
22. Graves A, Mohamed A-r, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada
23. Han M, Hua W, Xu W, Gong Y (2002) An integrated baseball digest system using maximum entropy method. In: *ACM International Conference on Multimedia*, Juan Les Pins, France
24. Hayes P (1996) *A catalog of temporal theories*. University of Illinois, Illinois



25. Hinton G, Deng L, Yu D, Dahl G, Mohamed A-r, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine* 29(6): 82–97
26. Hiroki I, Takiguchi T, Arika Y (2012) 3D tracking of soccer players using time-situation graph in monocular image sequence. In: *International Conference on Pattern Recognition (ICPR 2012)*, Tsukuba, Japan
27. Ibrahim ZAA (2007) *Characterisation des structures audiovisuelles par analyse statistique des relations temporelles*. University of Paul Sabatier, Toulouse
28. Ibrahim ZAA, Ferrane I, Joly P (2006) Conversation detection in audiovisual documents: temporal relation analysis and error handling. In *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Paris, France
29. Ibrahim ZAA, Ferrane I, Joly P (2011) A similarity-based approach for audiovisual document classification using temporal relation analysis. *EURASIP Journal on Image and Video Processing*, 2011
30. ISO-10744 (1992) *Information technology - hypermedia / time-based structuring language (HyTime)*. ANSI, New York
31. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014) Large-scale video classification with convolutional neural networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Washington, USA
32. Kautz H, Ladkin P (1991) Integrating metric and qualitative temporal reasoning. In: *AAAI-91*, California, USA
33. Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Nevada, USA
34. Krokhin A, Jeavons P, Jonsson P (2003) The tractable subalgebras of Allen's interval algebra. *ACM* 50(5): 591–640
35. Ladkin P (1986) Time representation: a taxonomy of interval relations. In: *National Conference on Artificial Intelligence*, Pennsylvania, USA
36. Ladkin P (1987) *The logic of time representation*. University of California, Berkeley
37. Ligozat G (1991) On generalized interval calculi. In: *National Conference on Artificial Intelligence (AAAI-91)*, California, USA
38. Ligozat G, Bestougeff H (1989) On relations between intervals. *Inf Process Lett* 34(4):177–182
39. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA
40. Meiri I (1996) Combining qualitative and quantitative constraints in temporal reasoning. *J Artif Intell* 87(1–2):295–342
41. Moulin B (1992) Conceptual graph approach for the representation of temporal information in discourse. *Knowl -Based Syst* 5(3):183–192
42. Navarette I, Marin R (1997) Qualitative temporal reasoning with points and durations. In: *International Joint Conference on Artificial Intelligence (IJCAI)*, Nagoya, Japan
43. Nebel B, Burckert H-J (1995) Reasoning about temporal relations: a maximal tractable subclass of Allen's interval algebra. *J ACM* 42(1):43–66
44. Pani AK, Bhattacharjee GP (2001) Temporal representation and reasoning in artificial intelligence: a review. *J Math Comput Model* 34(1–2):55–80
45. Petrovic M, Mihajlovic V, Jonker W, Djordjevic-Kajan S (2002) Multi-modal extraction of highlights from TV formula 1 programs. In: *IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland
46. Pujari A, Sattar A (1999) A new framework for reasoning about points, intervals and durations. In: *International Joint Conference on Artificial Intelligence (IJCAI)*, Stockholm, Sweden
47. Pujari A, Kumari V, Sattar, Abdul (1999) INDU: an interval and duration network. In: *Australian Joint Conference on Artificial Intelligence*, Australia
48. Qiu Z, Yao T, Mei T (2017) Deep quantization: encoding convolutional activations with deep generative model. In: *IEEE Conference on Computer Vision and Pattern Recognition*
49. Razavian AS, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Washington, USA
50. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA
51. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA
52. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39(6):1137–1149
53. Ross G (2015) Fast R-CNN. In: *IEEE International Conference on Computer Vision*, Santiago, Chile

54. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Zhiheng H, Karpathy A, Khosla A, Bernstein M, Berg A, Fei-Fei L (2015) Image net large scale visual recognition challenge. *Int J Comput Vis* 115(3): 211–252
55. Schwalb E, Vila L (1998) Temporal constraints: a survey. *Constraints* 3(2):129–149
56. Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. In: *International Conference on Neural Information Processing Systems*, Montreal, Canada
57. Tang S, Zhi M (2015) Summary generation method based on audio feature. In: *IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China
58. Tarski A (1941) On the calculus of relations. *Symbolic Logic* 6(3):73–89
59. Tavassolipour M, Karimian M, Kasaei S (2014) Event detection and summarization in soccer videos using Bayesian network and copula. *IEEE Trans Circ Syst Video Technol* 24(2):291–304
60. Tovinkere V, Qian RJ (2001) Detecting semantic events in soccer games: towards a complete solution. In: *IEEE International Conference on Multimedia and Expo (ICME2001)*, Tokyo, Japan
61. Van Beek P, Cohen R (1990) Exact and approximate reasoning about temporal relations. *Comput Intell* 6(3): 132–144
62. Vila L (1994) A survey on temporal reasoning in artificial intelligence. *J Artif Intel Commun* 7(1):4–28
63. Vilain MB (1982) A system for reasoning about time. In: *National Conference on Artificial Intelligence (AAAI82)*, Pittsburgh, USA
64. Vilain M, Kautz H (1986) Constraint propagation algorithms for temporal reasoning. In: *National Conference on Artificial Intelligence (AAAI86)*, Philadelphia, USA
65. Vilain M, Kautz H, Van Beek P (1990) Constraint propagation algorithms for temporal reasoning: a revised report. In: Weld DS, Klee JD (eds) *Readings in qualitative reasoning about physical systems*. Morgan Kaufmann, San Francisco, pp 373–381
66. Wang Q, Wan J, Yuan Y (2017) Deep metric learning for crowdedness regression. *IEEE Transactions on Circuits and Systems for Video Technology*
67. Q. Wang, J. Gao and Y. Yuan, "A Joint Convolutional Networks and context transfer for street scenes labeling," *IEEE Trans Intell Transp Syst*, vol. 19, no. 5, pp. 1457–1470, 2017
68. Wei L, Angelov D, Erhan D, Szegedy C, Reed S, Cheng-Yang F, Berg A (2016) SSD: single shot multibox detector. In: *European Conference on Computer Vision*, Amsterdam, Netherlands
69. Wetprasit R, Sattar A (1998) Temporal reasoning with qualitative and quantitative information about points and durations. In: *National Conference on Artificial Intelligence (AAAI)*, Madison, USA
70. Xie L, Chang S-F, Divakaran A, Sun H (2002) Structure analysis of soccer video with hidden markov models. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2002)*, Florida, USA
71. Yu H, Wang J, Huang Z, Yang Y, Xu W (2016) Video paragraph captioning using hierarchical recurrent neural networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
72. Zha S, Luisier F, Andrews W, Srivastava N, Salakhutdinov R (2015) Exploiting image-trained CNN architectures for unconstrained video classification. In: *British Machine Vision Conference*
73. Zhang S, Zhang C (2002) Propagating temporal relations of intervals by matrix. *Appl Artif Intell* 16(1):1–27
74. Zhang G, Liang G, Li W, Fang J, Wang J, Geng Y, Wang J-Y (2017) Learning convolutional ranking-score function by query preference regularization. In: *International Conference on Intelligent Data Engineering and Automated Learning*, Guilin, China
75. Zhou W, Vellaikal A, Kuo CCJ (2000) Rule-based video classification system for basketball video indexing. In: *ACM Workshops on Multimedia*, New York, USA