



Semantic Explicit Representation of Editing Patterns in Film

Ba Tu Truong

Institute for Multi-Sensor Processing & Content Analysis (IMPCA)
Department of Computing
Curtin University of Technology
Western Australia

Svetha Venkatesh

Institute for Multi-Sensor Processing & Content Analysis (IMPCA)
Department of Computing
Curtin University of Technology
Western Australia

In this paper, we present a graph-based visualization concept called the Double-Ring Take-Transition-Diagram (DR-TTD) that can capture and express the internal structure of a film scene and its editing patterns. The DR-TTD representation exhibits essential properties such as fully automatic construction, compactness, clarity, temporal perseverance and explicitly links to semantics. It presents takes and their transitions via nodes and edges of a 'graph' consisting of two rings as its backbone. All steps in the DR-TTD construction, including node classification, connecting nodes via edges and unit linking, are motivated from the understanding of film grammar for shot arrangement. In addition, there are signatures for filmic and semantic elements made explicit in this representation and these include dialogue, moving between zones/dramatic progression, shot association, introduction and resolution, master shot, non-dialogue narration and film editing orchestration.

Department of Computing, Curtin University of Technology, Perth, Western Australia

Draft Version 0.0 First time using IMPCA technical
report template

Contents

1	Introduction	1
2	Previous work	2
3	Double-Ring Take-Transition-Diagram	3
3.1	Node	3
3.1.1	Definitions	3
3.1.2	Representation	4
3.2	Edge	6
3.2.1	Definition	6
3.2.2	Representation	7
3.2.3	Mathematical Properties	7
3.3	Primitive Sequences of \mathcal{O} -nodes	8
3.4	Unit	9
3.5	DR-TTD Construction	9
4	Recognizing Film Semantics from DR-TTDs	11
4.1	Moving between Zones and Dramatic Progression	12
4.2	Shot Association	15
4.3	Scene Introduction and Resolution	17
4.4	Master Shot	19
4.5	Non-dialogue Narration & Familiar Images	20
4.6	Orchestration in Film Editing	22
5	Conclusions	24

1 Introduction

One problem facing current multimedia content management systems is the large gap between the rich meaning that users want when they query and browse media and the low level nature of content descriptions that we can actually compute. A serious need therefore exists to develop algorithms and technologies that can automatically annotate content and establish semantic connections between form and function, allowing users to access and navigate the indexed media in many interesting ways. Upon recognizing this problem, Dorai and Venkatesh [1] have proposed the *Computational Media Aesthetics* (CMA) framework for high-level content analysis of media. It is defined as “the algorithmic study of a variety of image and aural elements with insights from film grammar. It is also the computational analysis of the principles that have emerged underlying their manipulation of creative art of clarifying, intensifying and interpreting an event for audience.” As seen in Fig. 1, CMA advocates drawing guidance from media production principles, namely Film Grammar, for systematic analysis. It aims at offering the user high level semantics as intended by the filmmaker, both structural and expressive, in browsing, searching and navigating film and video documents.

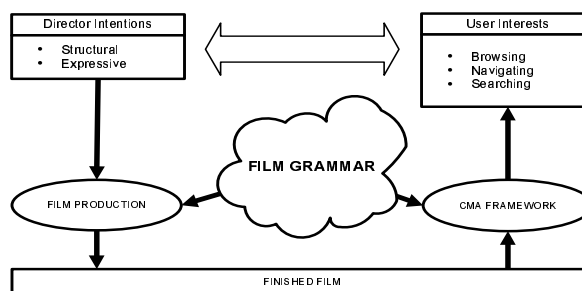


Figure 1: The CMA framework.

The aim of this work is to study the application of the CMA framework to discover semantics embedded in the editing patterns of a film scene. Instead of a fully automatic approach, we investigate how meanings can be uncovered from signature arrangements of film takes and their transitions in a Double-Ring Take-Transition-Diagram (DR-TTD) [2].

A film take is defined as “one uninterrupted run of the camera to expose a series of frames,” according to the Dictionary of Film Terms. A film take is also known as a shot captured during the film shooting¹ and before the editing stage, as opposed to shots in the finished film which are generally understood in multimedia research literature as the portion of the visual stream between two consecutive cut points in an edited film. As seen in Fig. 2, the filmmaker shoots many takes for a scene and during the film editing stage, different portions of selected takes are spliced together to produce the intended film scene. The term ‘take’ used in this work literally means a set of ‘edited’ shots that belong to the same production shot. We have 4 such takes

¹During film shooting, a (production) shot is a set of production takes and the notation “Shot X, Take Y” is used to distinguish between them.

in Fig. 2, although 5 takes are produced during the film shooting.

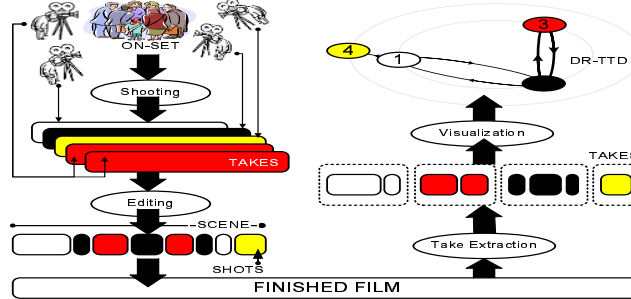


Figure 2: Takes, shots, scenes and DR-TTD.

Structurally, the take is the middle layer between the shot and the scene. This layer is rarely investigated but rich in semantics, as the arrangement of shots from takes shows the mediation level and narrative/dramatic intentions, in terms of editing, that the filmmaker applies on the film content. A signature arrangement in the DR-TTD exists for a given semantics because (a) the filmmaker relies on film grammar and film syntax, that guide how shots should be arranged, to create certain dramatic impact, or simply to avoid disorienting the viewer and (b) the construction of the DR-TTD is motivated from film grammar.

2 Previous work

An extracted take is essentially a cluster of ‘identical’ shots. Clustering of shots for the purpose of content browsing and presentation has been examined in [3] which clusters several visual features to create a hierarchical view of video content. Recently, we investigated the use of clustering to detect film scenes that are coherent in time/space or mood and have presented them in a Scene-Cluster Temporal Chart [4].

Shot clustering/grouping has often been used as an intermediate step in extracting scene boundaries [5, 6, 7]. These methods, therefore, do not demand that shots clustered together come from the same take, but from the same scene. They then use overlapping link reasoning to merge separated clusters into scenes. [6] proposes a technique called time-adaptive grouping to create a table-of-content for a video document. The authors attempt to incorporate shot length and shot activity into the shot similarity measure.

[8] studies the problem of mining video editing rules by performing row and column analysis on a matrix formed by shot indices and 3 shot attributes: distance, camera work and duration. [9] proposes a video editing support system that exploits film grammar related to shot size and camera work.

The DR-TTD visualization is based on the concept of Scene Transition Graph (STG) [5]. They both represent clusters and transitions among them via nodes and directed edges of a graph. However, a DR-TTD extensively exploits film grammar

to express richer semantics and its purpose is not to detect the scene transitions (by searching for cut edges in the graph), but to show the internal structure of a scene and make certain semantics explicit.

Shots, scenes and takes serve as the main input for our visualization process. The extraction of shot/scene indices is a well documented problem and many solutions are provided in the literature. Following is the summary of the four-steps in our take extraction technique described in [10]:

1. Check if the scene is action-driven or drama-driven by examining its tempo characteristics. Discard the scene if it is action-driven, as detecting takes for these kinds of scenes is difficult and less useful.
2. Compute the proximity matrix that measures the similarity between all pairs of shots in the scene.
3. Create shot groups by using a conventional clustering method and the proximity matrix.
4. Employ rules and conventions in film editing to merge and split clusters to further improve the results.

Our experimental results on 10 movies indicate that it is useful to divide the frame into sub-blocks and to measure shot similarity as the maximum of keyframe similarities. The performance is also better for dramatic-oriented, well edited films than action-based films [10].

3 Double-Ring Take-Transition-Diagram

In this section, we describe four important elements of a DR-TTD: node, edge, sequence and unit. Understanding these elements is essential to recognising the usefulness of a DR-TTD. The following notations are used: $\{x | \mathbf{Cond}(x)\}$ for the set of elements x , where condition $\mathbf{Cond}(x)$ is satisfied; and $|\mathbf{X}|$ for the number of elements in list/set/vector \mathbf{X} .

Let $\mathcal{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_n\}$ denote the shot sequence of a scene in temporal order, and $\mathbf{T} = \{\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_m\}$ denote the the set of takes extracted from this scene. Since each shot belongs to one and only one take, we have: $\mathbf{T}_i \subset \mathcal{S}$, $\mathbf{T}_i \cap \mathbf{T}_j = \emptyset$, and $\mathbf{T}_1 \cup \mathbf{T}_2 \cup \dots \cup \mathbf{T}_m = \mathcal{S}$.

3.1 Node

3.1.1 Definitions

Each take is represented by a node in the DR-TTD. The backbone of a DR-TTD consists of two circles where all nodes are placed. There are two kinds of nodes:

- *\mathcal{I} -nodes*. These nodes represent takes with at least two member shots. The set of all \mathcal{I} -nodes is denoted by \mathcal{I} . $\mathcal{I} = \{\mathbf{T}_i | \mathbf{T}_i \in \mathbf{T}, |\mathbf{T}_i| > 1\}$. Nodes of this kind are placed on the inner ring of the DR-TTD.

- *O-nodes*. These nodes represent takes with only one member shot. The set of all *O*-nodes is denoted by \mathcal{O} . $\mathcal{O} = \{\mathbf{T}_i \mid \mathbf{T}_i \in \mathbf{T}, |\mathbf{T}_i| = 1\}$. Nodes of this kind are placed on the outer ring of the DR-TTD.

This classification of nodes is core to the DR-TTD representation. It is motivated by the discovery that each node kind has its own distinctive set of semantic implications in film. Nodes on *O*-ring have functions that include adding drama, excitement, and highlights to the story. They tend to contain information that is less important for understanding the story, but help to maintain the continuity and logical flow of the narrative. Nodes on *I*-ring are important because they are repeated, emphasized and carry the major portion of the plot dialogue. They are vital in understanding the narrative development.

DR-TTD Property 1 *The semantic functions of *I*-nodes and *O*-nodes in a film scene are distinctive and fall into at least one category listed in Table 1.*

In addition, [?] describes eight cinesthetic elements that provide aesthetic gratification in film. Except for orchestration and parallel action, other cinesthetic elements are linked to the ring indices of the nodes. Separation, familiar image and master shot principle relate to nodes on the *I*-ring, as shots used to construct these elements are repeated throughout the scene. Slow disclosure, moving camera and multi-angularity are assembled from fragmented shots and relate to nodes on the *O*-ring.

Apart from being semantically motivated, the use of these two rings and node classification also simplifies many visualization aspects as evident later in this chapter.

3.1.2 Representation

Nodes are represented by circles and two smaller half circles are used to indicate if the take contains the first/last shot of the scene. It is useful to further incorporate the visual characteristics of a take into the appearance of its node. The first method computes the take features using every shot belonging to the take. The feature, e.g., average color, can be used to “label” the node.

An alternative method is to “label” the node with the whole image. It is more appropriate to extract only one frame from the shot sequence to represent the take rather than combine many images into one. The use of iconic images is useful as they can give a user a rough idea about the take angle, distance and its subjects. This method can proceed by first selecting a shot from all shots in the take and then selecting a representative frame (*R*-frame) from the *R*-frame list of the selected shot. The following three factors should be considered in selecting a shot:

- The distance from the shot to the centroid (\mathcal{C}) of all shots in the take. We approximate the distance by the average distance of a shot to other shots in the take. It is desirable to select the shot close to the centroid.

Table 1: Semantics for different node types.

	Description
\mathcal{O}-Nodes	
O_m	Movement into/out of/within dialogue. Quite often, as a result of movements, there is a change in the shot-size within the take.
O_p	Point of view/cut-away shots. This shot often follows the shot of a character looking off screen; it shows the objects as seen by this character.
O_c	Close-up of background objects or activity. When used at the beginning of the scene, it is a device for scene introduction. When used within the scene, it clarifies certain detail, e.g., letters written on a note, the gate sign, etc.
O_h	Highlight certain emotional expression of characters. It is often located within or just at the end a dialogue sequence, and has bigger shot size than shots making up the dialogue sequence, especially shots/takes of the same character.
O_v	Alternative view. This kind of shot is used to show characters and the setting from a different angle. It increases the depth and the sense of 3D space being projected.
O_l	Character looking towards a moving subject.
O_e	Establishing shot. Introducing the setting or a full shot of many characters that show their positions relative to each other.
\mathcal{I}-Nodes	
I_d	Main dialogue shots.
I_c	Centre shots, familiar images.
I_o	Interested on-looker of the main dialogue.

- The length of the shot (\mathcal{T}). It is important to select the shot that is most dominant for the take. Such shots are often indicated by their long duration.
- Motion level of the shot (\mathcal{M}). If the shot has a lot of motion, it may be a transitional shot and just not representative of the take. Hence, we prefer to select shots that have low motion.

Currently, we combine these factors linearly (after Gaussian-normalization) to select a shot to represent the take:

$$\mathcal{R}(\mathbf{T}_i) = (\mathbf{S}_{i_k} | 1 \leq k \leq t, \text{maximize}(-\mathbf{S}_{i_k}^{\mathcal{C}} + \mathbf{S}_{i_k}^{\mathcal{T}} - \mathbf{S}_{i_k}^{\mathcal{M}})),$$

where $\mathbf{S}_{i_1}, \mathbf{S}_{i_2}, \dots, \mathbf{S}_{i_t}$ are t shots of take \mathbf{T}_i .

For a selected shot \mathbf{S}_i , we then select the \mathcal{R} -frame accounting for the longest duration of the shot as its most representative \mathcal{R} -frame. That is,

$$\mathcal{R}(\mathbf{S}_i) = (\mathbf{F}_{i_k} | 1 \leq k < u, \text{maximize}(i_{k+1} - i_k)),$$

where $\mathbf{F}_{k_1}, \mathbf{F}_{k_2}, \dots, \mathbf{F}_{k_u}$ are u \mathcal{R} -frames of shot \mathbf{S}_i , with \mathbf{F}_{k_1} and \mathbf{F}_{k_u} being the first and last frame of this shot.

3.2 Edge

3.2.1 Definition

In a STG, two nodes \mathbf{T}_i and \mathbf{T}_j are connected by a directed edge if there is an index u such that \mathbf{S}_u in \mathbf{T}_i and \mathbf{S}_{u+1} is in \mathbf{T}_j . We extend this by using the width of each edge to indicate how much interaction occurs between the two takes. The interaction level indicates whether the two shots are loosely or strongly tied as a semantic unit. The width \mathcal{E} of an edge between two nodes \mathbf{T}_i and \mathbf{T}_j is calculated as:

$$\mathcal{E}(\mathbf{T}_i, \mathbf{T}_j) = |\{t | 1 \leq t \leq n - 1, \mathbf{S}_t \in \mathbf{T}_i, \mathbf{S}_{t+1} \in \mathbf{T}_j\}|$$

An edge between them is claimed if and only if $\mathcal{E}(\mathbf{T}_i, \mathbf{T}_j) > 0$. Let \mathbf{E} denote the set of all edges. \mathbf{E} is comprised of 4 subsets: $\mathbf{E}_{\mathcal{I}\mathcal{I}}, \mathbf{E}_{\mathcal{I}\mathcal{O}}, \mathbf{E}_{\mathcal{O}\mathcal{I}}$ and $\mathbf{E}_{\mathcal{O}\mathcal{O}}$, where \mathcal{I} and \mathcal{O} subscripts indicate the ring indices of the starting and ending nodes respectively. The following property can be deduced from nodes and their linking edges.

DR-TTD Property 2 *Two-way connected \mathcal{I} -nodes often show a dialogue between two characters. Each node focuses on one character (or a group characters) of the scene.*

The confidence level of this inference is proportional to the thickness of connecting edges. This inference lends itself to the fact that shot/reverse-shot, also known as shot/counter-shot, is the most commonly used technique for editing a dialogue sequence. Two alternating shots, generally in medium close-up or close-up, frame in turn the two speakers. Normally these shots are taken from the point of view of the person listening. But the spectator can assume the presence of both interlocutors even if the listener is not in the foreground because of the series of reverse angles. This type of editing follows the visual logic (character A is framed as she or he speaks - cut to character B framed as he or she speaks).

3.2.2 Representation

We represent different kinds of edges in a DR-TTD explicitly. An edge in $\mathbf{E}_{\mathcal{I}\mathcal{I}}$ is represented by an elliptical arc to show the interaction level of two \mathcal{I} -nodes. Circular arcs along \mathcal{O} -ring are used for $\mathbf{E}_{\mathcal{O}\mathcal{O}}$ -edges. $\mathbf{E}_{\mathcal{I}\mathcal{O}}$ -edges are represented by a straight line. If an $\mathbf{E}_{\mathcal{O}\mathcal{I}}$ edge connects an \mathcal{O} -node with its linked \mathcal{I} -node (a link within an unit, see Section 3.4), a straight line is used creating a double-headed edge, otherwise a circular arc is used (a link across units). In order to make sure that the arc does not cross either \mathcal{I} -ring or \mathcal{O} -ring, its centre is set on the straight line connecting the ring centre and the \mathcal{O} -node. For clarity, an $\mathcal{O}\mathcal{I}$ -edge should not intersect with the two rings. This can be achieved by having the center O' of its circular arc lie on the line passing through the ring center O and the \mathcal{O} -node.

3.2.3 Mathematical Properties

The DR-TTD representation is graph-based. However, due to the sequential nature of shot indices, not all weighted graphs are the DR-TTD representations of a certain film scene. There are some primitive mathematical properties regarding nodes and edges of a DR-TTD that can be used for validation:

- Since an \mathcal{O} -node contains only one shot, the total weight of all edges that start/end on an \mathcal{O} -ring node is not more than 1.

$$\forall \mathbf{T}_m \in \mathcal{O}, \sum_{\mathbf{T}_i \in \mathbf{T}} \mathcal{E}(\mathbf{T}_i, \mathbf{T}_m) \leq 1, \sum_{\mathbf{T}_i \in \mathbf{T}} \mathcal{E}(\mathbf{T}_m, \mathbf{T}_i) \leq 1$$

This also leads to the fact that the weight of any edge that starts/ends on the \mathcal{O} -ring is not more than 1.

$$\forall \mathbf{T}_i \in \mathcal{O}, \forall \mathbf{T}_j, \mathcal{E}(\mathbf{T}_i, \mathbf{T}_j) \leq 1, \mathcal{E}(\mathbf{T}_j, \mathbf{T}_i) \leq 1$$

The inequality occurs only at end nodes. We can also deduce that there is no more than one edge coming out/in an node on the \mathcal{O} -ring. This property allows us to construct a DR-TTD such that there are no crossings between any $\mathcal{O}\mathcal{O}$ -edge to any other edge (see Section 3.5).

- Since an \mathcal{I} -node contains more than one shot, the total weight of all edges that start (end) on a certain node on the \mathcal{I} -Ring is more than 1,

$$\forall \mathbf{T}_m \in \mathcal{I}, \sum_{\mathbf{T}_i \in \mathbf{T}} \mathcal{E}(\mathbf{T}_i, \mathbf{T}_m) \geq 1, \sum_{\mathbf{T}_i \in \mathbf{T}} \mathcal{E}(\mathbf{T}_m, \mathbf{T}_i) \geq 1$$

and, for the the equality to be obtained, the node must be an end node.

- Because shots are arranged as a linear sequence, the difference between the total weight of all edges coming in an arbitrary node \mathbf{T}_m and the total weight of all edges coming out of that node is not more than 1.

$$\left| \sum_{\mathbf{T}_i \in \mathbf{T}} \mathcal{E}(\mathbf{T}_i, \mathbf{T}_m) - \sum_{\mathbf{T}_i \in \mathbf{T}} \mathcal{E}(\mathbf{T}_m, \mathbf{T}_i) \right| \leq 1$$

The equality occurs only at end nodes.

3.3 Primitive Sequences of \mathcal{O} -nodes

A non-dialogue sequence is often constructed from the following two primitive structural elements (see Fig. 3). These elements can be deployed for the entire scene or combined together to construct the scene:

- **Montage sequence:** In montage sequences, the filmmaker assembles non-repeated shots from multiple angles/time/places to create a unified dramatic concept. Each take therefore contains only one shot and is placed consecutively on the \mathcal{O} -ring (see Fig. 3(a)).
- **Centre-shot/Cut-away sequence:** A video sequence is sometimes constructed by linking different actions to a central action. For example, in order to film a person looking around, the filmmaker may show the man's face repeatedly, following each shot by the shot of his current viewpoint. The DR-TTD of this sequence shows an \mathcal{I} -node connected to many \mathcal{O} -nodes (see Fig. 3(b)).

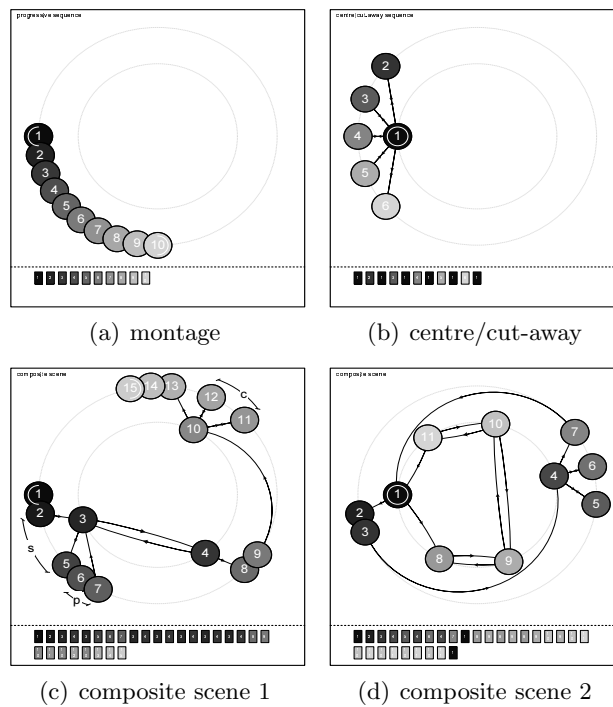


Figure 3: Scene constructs and DR-TTDs.

One important property of a DR-TTD is that every \mathcal{O} -node is part of either a montage sequence or centre sequence. In Fig. 3(c) and Fig. 3(d) which show two hypothetical scenes that use more than one of the above two sub-constructs, montage sequences are (1,2), (5,6,7), (8,9), (14,15) and (2,3) and the centre-shot sequences are (11,12,13) and (5,6,7). All \mathcal{O} -nodes in a given sequence are arranged in a anti-clockwise direction.

3.4 Unit

\mathcal{O} -node sequences can be linked backward directly or indirectly to \mathcal{I} -nodes. The exception is the first sequence that contains the starting shot, which can link forward to an \mathcal{I} -node. This linking forms different units in a DR-TTD. We often see that the content of an \mathcal{I} -node is related to the content of its linked sequences. In Fig. 3(c), the links are from sequence (1,2) (forward) to node 3, (5,6,7) to node 3, (8,9) to node 4, (11,12,13) to node 10 and (14,15) indirectly to node 10, whilst in Fig. 3(d), the links are from sequence (2,3) to node 1 and (5,6,7) to node 4. In this way, each \mathcal{O} -node sequence can be assigned to one and only one \mathcal{I} -node. For each \mathcal{I} -node, all its associated sequences are arranged in an anti-clockwise direction; the ordering also reflects their temporal order in the video sequence. An \mathcal{I} -node not linked to any \mathcal{O} -node forms a singleton unit, e.g., \mathcal{I} -nodes 8,9,10,11 in Fig. 3(d). In Figures 3(c) and 3(d), we have totals of 3 and 6 units respectively.

In terms of graph representation, unit linking allows the arrangement of \mathcal{O} -node sequences on the \mathcal{O} -ring to depend entirely on the arrangement of \mathcal{I} -nodes on the \mathcal{I} -ring, which simplifies the graph construction. In terms of semantics, this linking, as shown in the property below, is motivated from the association between the contents of \mathcal{I} -nodes and its linked \mathcal{O} -node sequences. This property also relates to the functions of \mathcal{O} -nodes and \mathcal{I} -nodes described in DR-TTD Property 1 (see Table 1).

DR-TTD Property 3 *The grouping of \mathcal{O} -nodes to \mathcal{I} -nodes are semantically meaningful and the semantic associations are listed in Table 2.*

3.5 DR-TTD Construction

The most important problem in constructing a DR-TTD is the exact placement of \mathcal{I} -nodes and \mathcal{O} -nodes. Our objectives are: (a) *clarity*, by minimizing the number of edge crossings, an important aspect of most graph layout algorithms; (b) *consistency*, by appropriate spacing; (c) *preserving temporal order*, by arranging takes in a left-to-right anti-clockwise direction to the extent possible; and (d) *preserving sub-constructs* by placing their elements close and appropriately spaced on the two rings. We describe below five steps that will automatically construct the DR-TTD in line with the specified objectives:

STEP 1: Numbering takes: Each take is numbered according to the order of the first shot in the take. Hence, for $\mathbf{T}_i = \{\mathbf{S}_{i_1}, \mathbf{S}_{i_2}, \dots, \mathbf{S}_{i_m}\}$ and $\mathbf{T}_j = \{\mathbf{S}_{j_1}, \mathbf{S}_{j_2}, \dots, \mathbf{S}_{j_m}\}$, we have $i < j \iff i_1 < j_1$. This numbering reflects the ordering of takes in the shot sequence.

STEP 2: Forming \mathcal{O} -node sequences (see Section 3.3) and assigning \mathcal{O} -node sequences to \mathcal{I} -nodes (see Section 3.4).

The remaining steps focus on how to select an ordering of \mathcal{I} -nodes. Let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_{|\mathcal{I}|}\}$, $\mathbf{T}_{\alpha_i} \in \mathcal{I}$ be an ordering of \mathcal{I} -nodes from the left to right in an anti-clockwise direction, and α_1 is always the \mathcal{I} -node with the lowest index. Let A_0 is the set of all possible ordering of \mathcal{I} -nodes, we have $|A_0| = (|\mathcal{I}| - 1)!$. Let

Table 2: Semantics of grouping \mathcal{O} -nodes to \mathcal{I} -nodes.

	Description (Assume \mathcal{X} is the set of \mathcal{O} -nodes linked to \mathcal{I} -node \mathcal{Y})
U_l	<i>leading to.</i> All shots in \mathcal{X} logically lead to the existence of \mathcal{Y} . This often occurs at the start of the scene. This is related to \mathcal{O} -node function O_m, O_e .
U_r	<i>reacting to.</i> Shots in \mathcal{X} show the reaction of other characters to the action of a character in the shot of \mathcal{Y} that precede \mathcal{X} . This is related to \mathcal{O} -node function O_h .
U_o	<i>continuing from.</i> Shots in \mathcal{X} show the continuity of action of the character shown in the shot \mathcal{S} of \mathcal{Y} that precede \mathcal{X} . For example, shot \mathcal{S} of \mathcal{Y} shows a character raising from a chair, all shots in \mathcal{X} show the character walking away. This is related to \mathcal{O} -node function O_m .
U_c	<i>centering around.</i> Action in all \mathcal{X} shots center around the action in \mathcal{Y} . This is related to centre/cut-away shots described in the previous section.

$\mathcal{C}(\alpha, \mathbf{E}_1, \mathbf{E}_2)$ be the number of crossings between an edge in \mathbf{E}_1 to a different edge in \mathbf{E}_2 . Due to the chosen structure and placements, we have: $\mathcal{C}(\alpha, \mathbf{E}_{\mathcal{O}\mathcal{O}}, \mathbf{E}) = \mathcal{C}(\alpha, \mathbf{E}_{\mathcal{I}\mathcal{I}}, \mathbf{E}_{\mathcal{I}\mathcal{O}}) = \mathcal{C}(\alpha, \mathbf{E}_{\mathcal{I}\mathcal{I}}, \mathbf{E}_{\mathcal{O}\mathcal{I}}) = \mathcal{C}(\alpha, \mathbf{E}_{\mathcal{I}\mathcal{O}}, \mathbf{E}_{\mathcal{I}\mathcal{O}}) = 0$. Thus, we only need to deal with $\mathcal{C}(\alpha, \mathbf{E}_{\mathcal{I}\mathcal{I}}, \mathbf{E}_{\mathcal{I}\mathcal{I}})$, $\mathcal{C}(\alpha, \mathbf{E}_{\mathcal{I}\mathcal{O}}, \mathbf{E}_{\mathcal{O}\mathcal{I}})$ and $\mathcal{C}(\alpha, \mathbf{E}_{\mathcal{O}\mathcal{I}}, \mathbf{E}_{\mathcal{O}\mathcal{I}})$.

STEP 3: Minimizing crossings in $\mathcal{I}\mathcal{I}$ -Edges: The number of crossings between two $\mathcal{I}\mathcal{I}$ -edges depends only on their ordering and not their actual placement, since they are placed in a circle. Let A_1 denote the set of all arrangements in a left-to-right, anti-clockwise direction. We have:

$$A_1 = \{\alpha | \alpha \in A_0, \text{minimize}(\mathcal{C}(\alpha, \mathbf{E}_{\mathcal{I}\mathcal{I}}, \mathbf{E}_{\mathcal{I}\mathcal{I}}))\}$$

Apart from ensuring that the graph has a clear layout, minimizing the number of crossings among $\mathcal{I}\mathcal{I}$ -edges semantically places strongly tied nodes close to each other. Note that if \mathcal{I} is comprised of connected components $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_k$ then we need to search in $|\mathcal{I}_1|! + |\mathcal{I}_2|! + \dots + |\mathcal{I}_k|!$ ($\leq |A_0|$) orderings to find A_1 .

STEP 4: Minimizing crossings in \mathcal{O} -related-Edges: For each \mathcal{I} -node ordering in A_1 , we can calculate the placement of all nodes. The calculation is based on equally spaced units and three spacing parameters (p, c, s) for spacing between two progressive nodes, two \mathcal{O} -nodes of the same centre sequence and two \mathcal{O} -node sequences linked to the same \mathcal{I} -node (see Fig. 3(c)). Unit boundaries (7,8), (9,11) and (15,1) in Fig. 3(c) are equally spaced, and so are (3,8), (8,9), (9,5), (7,10), (10,11), and

(11,2) in Fig. 3(d). The three spacing parameters often take on default values but can be adjusted automatically to accommodate larger number of nodes. After calculating the placement of all nodes, we search for arrangements with the minimum number of crossings of edges that are not \mathcal{II} -edges to further ensure the clarity of the graph. Then,

$$A_2 = \{\alpha | \alpha \in A_1, \text{minimize}(\mathcal{C}(\alpha, \mathbf{E}_{\mathcal{IO}}, \mathbf{E}_{\mathcal{OI}}) + \mathcal{C}(\alpha, \mathbf{E}_{\mathcal{OI}}, \mathbf{E}_{\mathcal{OI}}))\}$$

STEP 5: Preserving temporal order of \mathcal{I} -Nodes: The last step of the algorithm aims at preserving as much as possible the temporal ordering of \mathcal{I} -nodes. This is done by maximizing the number of pairs of \mathcal{I} -nodes that are in the right order. Let A_3 denote a set of such arrangements. Then,

$$A_3 = \{\alpha | \alpha \in A_2, \text{maximize}(\sum_{j>i} \text{sign}(\alpha_j - \alpha_i))\}$$

Now, we just select one ordering of \mathcal{I} -nodes from A_3 . Note that if there is no \mathcal{I} -node, we can stop after STEP 1 and arrange all nodes on the \mathcal{O} -ring.

4 Recognizing Film Semantics from DR-TTDs

So far we have described three properties that are inherent in the DR-TTD construction:

1. Semantic distinction between \mathcal{O} -nodes and \mathcal{I} -nodes.
2. Dialogue from \mathcal{II} -edges.
3. Semantic associations in unit groupings.

Here we will describe six other properties of the DR-TTD that are concerned with the association between semantics and signature arrangement of nodes and edges:

4. Moving between zones and dramatic progression.
5. Shot association.
6. Scene introduction and resolution through \mathcal{O} -node sequence.
7. Master shot configuration.
8. Non-dialogue narration and familiar images.
9. Editing orchestration.

These properties are discovered based on the understanding of film grammar and detailed examination of the DR-TTD representation of hundreds of scenes from several movies that include American Beauty (AB), Erin Brockovich (EB), The

Matrix (MX), The Truman Show (TS), The Siege (SG), The Mummy (MM), Sleepy Hollow (SH).

This section is organized into six sub-sections, and each section describes in detail one of the six properties (4-9 in the above listing). At least four examples of real film scenes are used in each section to verify the proposed property. In addition, all three properties described previously (2-3 in the above listing) shall be verified in each sub-section using these examples. Interested readers may want to refer to [11] for the detailed verification of DR-TTD Property 1, which shows most \mathcal{I} -nodes and \mathcal{O} -nodes in examples described here have distinctive semantic functions as listed in Table 1.

Table 3 lists additional notation to be used in the section.

Table 3: Notation used in Section 4.

$[\text{MM:TT}]$	A scene occurring in movie MM at time (minute) TT.
(x, y)	The pair of nodes x and y from the graph.
$(a, b, c) \mapsto y$	The linking of \mathcal{O} -nodes a, b, c to \mathcal{I} -node y .
$(m - n) \mapsto y$	The linking of \mathcal{O} -nodes $\{m, m + 1, \dots, n\}$ to \mathcal{I} -node y .
$m - n$	The edge between two nodes m and n .

4.1 Moving between Zones and Dramatic Progression

DR-TTD Property 4 *Moving between zones and the dramatic progression of a scene is evident from cut edges, especially \mathcal{II} -edges and \mathcal{OI} -edges.*

The location of a scene may be divided into smaller zones. During the scene narrative, characters may move from zone to zone or the dramatic focus may change to a different group of characters in the scene. For the purpose of content analysis, navigation and authorization, it is useful to make these zone transitions explicit, because the narrative portion in each zone is often self-contained and can be treated separately. Filmmakers have different camera set-ups for different zones, resulting in different take sets. Therefore, we have a cut edge² in DR-TTD when zone-to-zone transitions are made. We are primarily interested in cut edges that are also \mathcal{II} -edges or \mathcal{OI} -edges, since they connect different units of the scene as formed in Section 3.4. On the other hand, \mathcal{IO} -edges and \mathcal{OO} -edges always connect nodes within a unit. In addition, when a cut edge is an \mathcal{OI} -edge, we can generally conclude that the montage sequence linked to the starting node shows the transition between zones. Otherwise, the movement is often indicated in the last \mathcal{I} -node of the first zone. This

²An edge of a graph is a cut-edge if its deletion increases the number of components in the graph

is similar, but in finer granularity, to the use of cut edges in the STG for detecting scene changes [5].

Fig. 4(a) shows a scene in *The Thirteenth Floor* where detective Larry McBain meets Jason ([TF:16']). They first talk outside, and then move into the computer lab resulting in cut edge 5—6. The zone movement is shown in Take 5. The party scene of *American Beauty* is shown in Fig. 4(b) ([AB:29']). Two zones, one showing the conversation between Lester, Carolyn and Buddy and the other showing the conversation between Lester and Ricky, are separated by cut edge 4—5. The last shot of Take 4 sees Lester leaving the conversation to get a drink. In this scene, Take 9 also sets a new zone in which Buddy and Carolyn are talking at a table. However, the filmmaker uses a very long shot to cover it and the DR-TTD mistakenly groups it into the same unit as the conversation between Lester and Ricky. The cut \mathcal{IO} -edge is hence mistakenly ignored in our framework. If that dialogue is constructed via shot-reverse-shots then a cut \mathcal{II} -edge would have resulted and been recognized.

It should be noted that a cut edge does not occur in a scene involving movement between zones if the characters later return to the previous zone, or the camera follows the characters to the new zone and continues shooting the scene and/or combines it with new takes set up for the new zone.

The zone change is not the only narrative event that triggers a cut edge. It may also arise from the dramatic progression of the scene. The most common device for manipulating the dramatic emphasis is varying shot size (also called shot distance). The filmmaker may start all shots in medium, then later change to close-ups to indicate that the drama has heightened. Also, a piece of dialogue that is vital for advancing the plot requires close-ups or some shift in the pattern of shots to alert us that what we are hearing is more important than what we have heard earlier in the sequence [?].

For example, the dinner scene between Carolyn and Buddy shown in Fig. 4(c) ([AB:51']) has a cut edge 3—4 separating medium shots and close-ups. The close-ups indicate the increased intimacy between them towards the end of the scene. The zone change and dramatic progression may occur together in one single scene as seen in Fig. 4(d) and 4(e). Fig. 4(d) ([EB:28']) shows a scene in *Erin Brockovich* in which Erin first meets Donna Jensen at the door, and then moves to the lounge room, causing cut edge 3—4. As the conversation becomes more dramatic towards the end, close-ups shots are used and are separated from early medium shots via cut edge 4—6. Fig. 4(e) ([EB:114']) shows a similar scene between these two characters occurring later in the film, the meeting outside is separated with the conversation inside via cut edge 7—9, while cut edge 9—12 indicates the increased drama in their conversation.

Note that if the emotion changes in only one character, the DR-TTD would contain a triangle with a missing edge. Fig. 4(f) ([EB:35']) shows one such scene in *Erin Brockovich* when Erin talks to Ed about being fired. Ed (Take 2) is calm for the entire scene while Erin's emotion is intensified through the shifting from medium shots (Take 1) to close-ups (Take 3). This signature will be further explored in the next section.

Our detailed examination [11] shows that most of the \mathcal{I} -nodes and \mathcal{O} -nodes operate according to semantic functions listed in Table 1. The function of Take 7 in [TF:16']

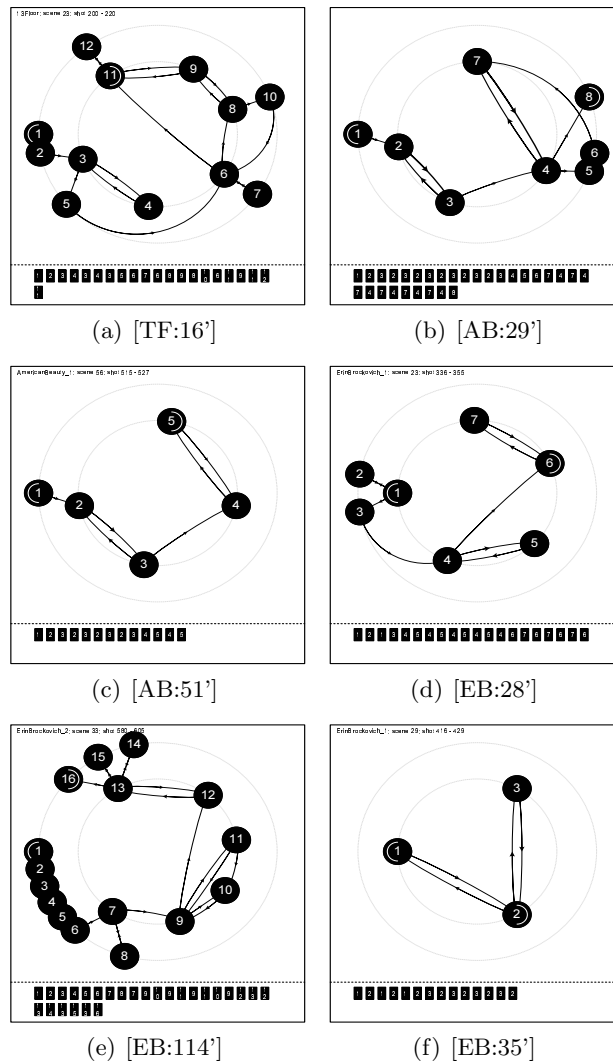


Figure 4: Zone/drama change examples.

is unclear; it is a shot in which the focus changes from Jason to Detective McBain. Take 12 is different in shot angle to Take 9, but not significant enough to suggest an alternative view (O_v). The closest semantic function is O_h , i.e., Detective McBain reacting to what Jason has said in previous shot. Take 9 of [AB:29'], as mentioned above, should be classified as the main dialogue component. Take 2 of [AB:51'] is Donna's brief speech. Takes 1 and 2 of [EB:28'] also contain brief dialogues.

Table 4 shows the verification of DR-TTD Property 2 and DR-TTD Property 3. They show that most sequence and unit groupings are meaningful. The groupings of U_l nature are $(1,2) \mapsto 3$, in [TF:16'], $1 \mapsto 2$ in [AB:51'], $(1-6) \mapsto 7$ in [EB:114']. The only grouping of U_r kind is $12 \mapsto 11$ in [TF:16']. Editing continuity is realised in four grouping instances, for example, $(7 \mapsto 6)$ in [EB:28'] is the shot of keyboard following the shot of Jason looking down. There is no clear association between \mathcal{I} -nodes and

Table 4: Verification of unit grouping and dialogue for examples in Fig. 4.

	[TF:16']	[AB:29']	[AB:51']	[EB:28']	[EB:114']
U_i	(1-2)→3	1→2	1→2		(1-6)→7
U_r	12→11				
U_o	10→8 7→6	(5,6)→4		2→1	8→7
U_c					(14-16)→13
GrpError	5→3	8→4		3→1	3→1
Dlg.	(3,4),(8,9) (9,11)	(2,3),(4,7)	(2,3),(4,5)	(4,5),(6,7)	(9,11) (12,13)
DlgError					(10-9)

\mathcal{O} -nodes for the following groupings: 9→5 in [TF:16'], 8→4 in [EB:28'], 3→1 in [EB:28'] and 3→1 in [EB:28']

The last two rows of the table show that almost all two-way connected \mathcal{I} -nodes are part of dialogue sequences. The only error the interpretation is for (10,9) in [EB:114']; George serves as an on-looker and does not participate in the dialogue.

4.2 Shot Association

DR-TTD Property 5 *If we have three \mathcal{I} -nodes heavily connected on two sides, with no edges on the remaining side, we can generally assume that the two unconnected \mathcal{I} -nodes are takes of the same character at different camera distance/angles.*

As an edge indicates the interaction between two nodes, its absence may be meaningful. If two nodes are linked via another node, we say they are being ‘indirectly’ linked. The indirect link is a device for changing from one shot to another shot in which the same character is captured with a different camera setup, as a direct link between two such nodes may result in discontinuity. Also, in dialogue sequences, dramatic emphasis may vary for only one character through different camera setups, while others remain the same (e.g., Fig. 4(f)). Therefore, if we have three \mathcal{I} -nodes heavily connected on two sides, with no edges on the remaining side, we can generally assume that the two unconnected \mathcal{I} -nodes are takes of the same character at different camera distance/angles.

For example, the missing link 2—3 in an *American Beauty* scene showing the Burnham family going to work (Fig. 5(a), [AB:3']) indicates two takes of the house pass-way at long (Take 2) and medium-long shots (Take 3). The latter allows a closer view of character’s faces and actions. Fig. 5(b) ([SH:5']) shows the courtroom scene in *Sleepy Hollow*; missing edge 5—6 relates to Takes 5 and 6 being the medium and close-up shots of Constable Crane. Missing edges 4—9, 4—10 and 9—10 relate to Takes 4, 9 and 10 being the medium, profile medium and medium close-up shots of the judge. In *The Matrix*, when Neo meets Trinity at the night club (Fig. 5(c), [MX:9']), Takes 5, 6 and 8 are shots of Trinity that relate to the missing edges. Fig. 5(d) ([AB:25']) shows the scene in *American Beauty* in which Ricky meets Jane at school. Missing edges 3—6, 3—7 and 6—7 link to Takes 3, 6 and 7 being long,

Table 5: Verification of unit grouping and dialogue for examples in Fig. 5.

	[AB:3']	[SH:5']	[MX:9']	[AB:25']
U _l		(1-3)→4	(1,2)→3	(9,10)→2
U _r	4→3	(7,8)→6,11→10		
U _o			7→6	
U _c				
Errors			(9,10)→8	
Dlg.	(1,2),(1,3)	(4,5),(6,9) (6,10)	(2,4),(3,4) (3,7),(3,5)	(1,2),(4,5),(2,6) (2,7),(7,8)
DlgError				(2,3)

The only error in dialogue interpretation occurs in Takes (2,3) in [AB:25'], where the existence of the shot-reverse-shot pattern is not due to dialogue, but from Jane and Angela looking at Ricky from a distance.

4.3 Scene Introduction and Resolution

DR-TTD Property 6 *The existence of montage sequences at the start and/or the end of a DR-TTD indicates that the filmmaker has visually staged the introduction and/or resolution to the main narrative in the scene.*

The introduction to a scene is achieved by showing:

- location and background activity.
- relative positions between characters.
- the current action of a character in close-up.

The resolution sequence at the end of a scene often provides:

- links to actions in the next scene.
- gradual conclusion of the scene.

For example, Fig. 6(a) shows the scene where Ricky talks to his dad in the car. The first three shots introduce the scene by showing the street, a hand over a notebook, and a shot of both Ricky and his dad. The resolution shot (Take 6) reveals what is written in Ricky's book, indicating the fact that he is a drug dealer. The court hearing scene in *Erin Brockovich* (Fig. 6(b), [EB:76']) shows a gradual approach to the main event. The first five shots show the parking lot, gate sign, hallway and wide shots of the whole court room from the back and front. Fig. 6(c) ([SH:12']) shows the scene in *Sleepy Hollow* where Constable Crane arrives at Sleepy Hollow. There are various shots (Takes 1-7) of party activities preceding the meeting between Constable Crane and the Van Tassel family (Takes 8, 11, 12, 13). The last two shots introduce new characters into the story and connect to the next scene. The last conversation between Lester and Angela (Fig. 6(d), [AB:104']) is concluded

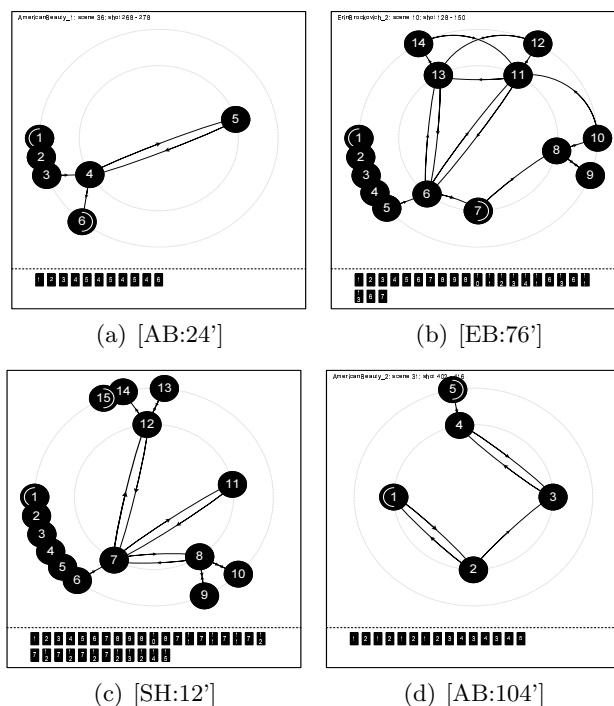


Figure 6: Introduction & resolution examples.

by Lester walking away (Take 5), linking to the next scene in which he is sitting in a room.

The absence of introduction sequences indicates that the filmmaker would like to have the spectator primarily focusing on the action (e.g., the conversation between Lester and Angela in Fig. 6(d)), and other details are provided later in the scene. Also, the scene may not resolve visually but through dialogue or characters' expressions. For example, Erin smiles at the outcome of the hearing (Take 7, Fig. 6(b)).

The verification of DR-TTD Property 1 in [11] shows that semantic functions of most \mathcal{O} -nodes and \mathcal{I} -nodes agree with DR-TTD Property 1. Table 6 shows the verification of unit grouping and dialogue. Most unit groupings are meaningful, conforming to DR-TTD Property 2. The 'lead to' association is evident in [AB:24'], [EB:76'] and [SH:12']. The continuity editing is found in $6 \mapsto 4$ in [AB:24']. The close-up shot of the book (Take 6) follows the shot of Ricky looking down (Take 4). Also, the last shot of Take 4 in [AB:104'] shows Lester rising from the chair, and Take 5 shows him continuing that movement.

The only error in dialogue inference (DR-TTD Property 3) occurs in scene [EB:76'] where characters exchange looks in a brief shot reverse-shot sequence.

Table 6: Verification of unit grouping and dialogue for examples in Fig. 6.

	[AB:24']	[EB:76']	[SH:12']	[AB:104']
U_l	(1-3) \mapsto 4	(1-5) \mapsto 6	(1-6) \mapsto 7	
U_r		14 \mapsto 13		
U_o	6 \mapsto 4	12 \mapsto 11		5 \mapsto 4
U_c		(9,10) \mapsto 8	(9,10) \mapsto 8,(13-15) \mapsto 12	
GrpError		12 \mapsto 11		
Dlg.	(4,5)	(6,13)	(7,8),(7,11), (7,12)	(1,2),(3,4)
DlgError		(6,11)		

4.4 Master Shot

DR-TTD Property 7 *In the DR-TTD representation, a master shot is indicated by an \mathcal{I} -node linked with many takes of shot-reverse-shot configurations.*

In classical Hollywood style, a master shot is a filmic recording of an entire scene, from start to finish, and taken from an angle that keeps all the players in view. It is ordinarily supplemented with other shots such as close-ups of individuals. It establishes an objective and stable perspective on a given situation [?]. The master shot gives a broad view and occurs at least twice in the course of a scene. The master shot is often interwoven with shot-reverse-shot sequences and the transition between a master shot to close shots of any individual character is natural and not disorientating. Therefore, in a DR-TTD a master shot often links with many takes of shot-reverse-shot configurations.

For example, in Fig. 7(a) ([SH:7']), Take 2 of the scene in *Sleepy Hollow* which shows Constable Crane performing an autopsy is a master shot which shows the table and all characters. It links with shot-reverse-shot sequences 1—3 and 6—7. Fig. 7(b) shows the scene outside the basketball court in *American Beauty* with Take 2 being the master shot showing all characters talking in a group. It is also linked with other shots in shot-reverse-shot sequences. Similarly, the master shot of the scene in which the Fitts family is having breakfast (Fig. 7(c), [AB:23']) is Take 4.

Many master shots may occur within a single scene as shown by Take 1 and Take 5 in the dinner scene of the Burnham family in *American Beauty* (Fig. 7(d)). These master shots capture the dinner table in long and medium long shots. The conclusion of a take being a master shot is reinforced if it occurs at the start/end of the scene as seen in [SH:7'], [AB:23'] and [AB:62'].

The verification of DR-TTD Property 1 in [11] shows the semantic functions of \mathcal{I} -nodes and \mathcal{O} -nodes agree well with Table 1. Table 7 shows the verification of DR-TTD Property 2 and DR-TTD Property 3. It is obvious from the above description that the groupings 1 \mapsto 2 in [AB:16'] and [AB:23'] are of the 'leading to' kind. The remaining grouping of \mathcal{O} -nodes to \mathcal{I} -nodes is classified as U_o . The close-up of the neck continues from the hand moving over the neck, 4 \mapsto 3 in [SH:7'].

Master shots introduce some errors into the dialogue inference. The two-way connected node pair (4,5) in [AB:23'] is not caused by any shot-reverse-shot sequence, and neither are (2,5) and (4,5) in [AB:62']. There are no actual dialogues

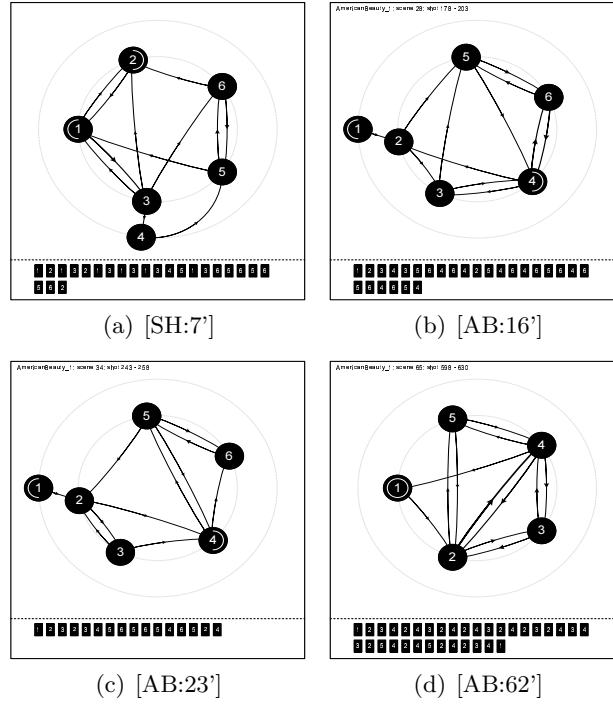


Figure 7: Master shot examples.

Table 7: Verification of unit grouping and dialogue for examples in Fig. 7.

	[SH:7']	[AB:16']	[AB:23']	[AB:62']
U_l				
U_r				
U_o	4 \rightarrow 3	1 \rightarrow 2	1 \rightarrow 2	
U_c				
GrpError				
Dlg.	(1,3),(1,2),(5,6)	(3,4),(4,6),(5,6)	(2,3),(5,6)	(3,4)
DlgError			(4,5)	(2,3),(2,5),(4,5)

between characters in (2,3) in [AB:62'] as Jane serves mainly as the onlooker of the conversation between Lester and Carolyn.

4.5 Non-dialogue Narration & Familiar Images

DR-TTD Property 8 *As an alternative to dialogue-based scenes evident in two-way linked \mathcal{I} -nodes, a non-dialogue narration is achieved by the use of montage and cut-away sequences.*

Our analysis so far has mainly focused on dialogue-based scenes. These scenes are characterized by dominant shot-reverse-shot sequences (two \mathcal{I} -nodes linked by two thick edges in DR-TTDs). However, drama-driven scenes do not necessarily

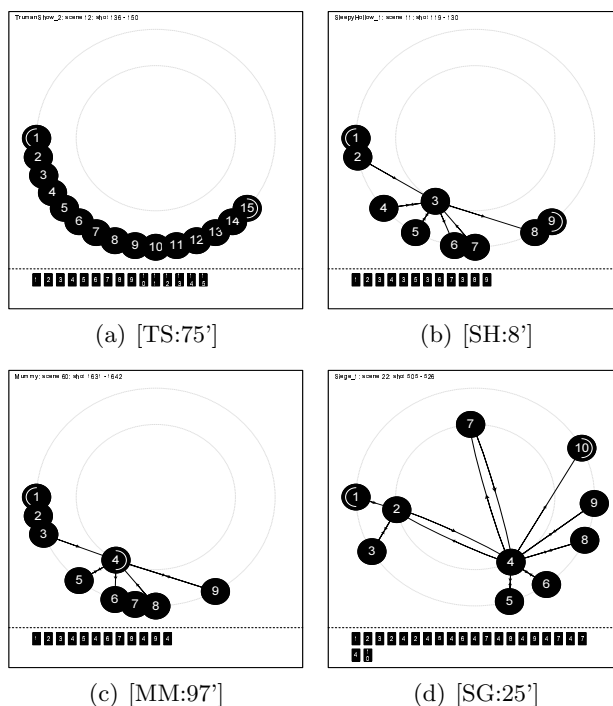


Figure 8: Non-dialogue narration examples.

contain dialogues. One way to construct such scenes is to use a series of progressing shots (montage sequence). These shots, each of which shows a fragmented part of the scene, are unified by the drama. For example, in order to create the scene where people are looking for Truman in *Truman Show* (Fig. 8(a), [TS:75']) the filmmaker uses crowd shots of different street corners at different angles and distances.

Any picture that reappears in a film with approximately the same framing and composition is called a familiar image. It is one of the devices for linking fragmented shots of a scene. The familiar image plays the role of a pivotal image around which a scene or part of a scene is constructed [?]. In the DR-TTD, familiar images show up as centre shots. For example, the scene where Constable Crane travels to the town in *Sleepy Hollow* is constructed through this device (Fig. 8(b), [SH:8']). The medium close-up shot of Constable Crane sitting in the horse carriage (Take 3) links all images of the road, forest, etc. Fig. 8(c) ([MM:97']) shows a scene in *The Mummy* where High Priest Imhotep tries to use Evelyn to resurrect his lover, and the shot of Evelyn tied on a table (Take 4) is used as the familiar image. At a briefing in *The Siege*, shown in Fig. 8(d) ([SG:25']), the shot of Agent Hubbard (Take 4) is the centre of action that links various shots of people listening.

As detailed in [11], there are uncertainties associated with the semantic functions of many \mathcal{O} -nodes, because these nodes do not function by themselves, but need to interact with one another to convey meaning.

The verification of DR-TTD Property 2 and DR-TTD Property 3 is shown in Table 8. The groupings of \mathcal{I} -nodes to \mathcal{O} -nodes in most examples are meaningful.

Table 8: Verification of unit grouping and dialogue for examples in Fig. 8.

	[TS:75']	[SH:8']	[MM:97']	[SG:25']
U_l		$(1,2) \mapsto 3$	$(1-3) \mapsto 4$	$(1 \mapsto 2)$
U_r				
U_o		$4 \mapsto 3, 9 \mapsto 3$		
U_c		$(1-2, 4-9) \mapsto 3$		$(5, 6, 8-10) \mapsto 4$
GrpError				$3 \mapsto 2$
Dlg.				$(4, 7)$
DlgError				$(2, 4)$

As usual, U_l kind of grouping is associated with starting shots, such as $(1,2) \mapsto 3$ in [SH:8'], $(1-3) \mapsto 4$ in [MM:97'], and $1 \mapsto 2$ in [SG:25']. All other takes in [SH:8'] center around the image of Constable Crane in Take 4, while most of the takes in [SG:25'] centres around the Agent Hubbard shot. The groupings $4 \mapsto 3$ and $9 \mapsto 3$ follow a U_o relation as they show what Constable Crane is looking at. Unfortunately, we can not find any clear semantic association in the grouping $3 \mapsto 2$ in [SG:25'].

There is one error in dialogue interpretation in (2,4) in [SG:25'], as there are no actual dialogues in the short shot-reverse-shot sequence that causes the two-way connected pair.

4.6 Orchestration in Film Editing

DR-TTD Property 9 *Orchestration in film editing is sometimes realized in the symmetry of the DR-TTD.*

[?] defines orchestration as the arrangement of various elements of structure throughout a scene or entire film, which includes the symmetry in the editing pattern for a film scene. This can be quickly recognized through the symmetry of the DR-TTD itself.

For example, Fig. 9(c) shows a simple scene (dinner between Agent Hubbard and Elise) in *The Siege* symmetric via shot-reverse-shot (Takes 2,3) pattern. Another scene in *The Siege* showing the meeting between Elise, Agent Hubbard and General Devereaux (Fig. 9(d), [SG:44']) also has a symmetric editing pattern around Take 3, the medium shot of Agent Hubbard. Fig. 9(a) ([EB:66']) shows the phone conversation between Erin and George in *Erin Brockovich*. Each take is used with rhythmic frequency and creates the symmetry in its DR-TTD. The scene in *American Beauty* when Lester and Ricky are smoking pot (Fig. 9(b), [AB:32']) has a symmetry around Take 2, a two-shot take of Ricky and Lester. It also starts and ends with the same take.

With respect to DR-TTD Property 1, all \mathcal{I} -nodes and \mathcal{O} -nodes of examples examined in this section operate according to the semantic functions listed in Table 1 [11].

Table 9 shows the verification of DR-TTD Property 2 and DR-TTD Property 3. It can be seen that the only two \mathcal{O} -node and \mathcal{I} -node linkings in these examples are both semantically meaningful as specified in DR-TTD Property 3. The first

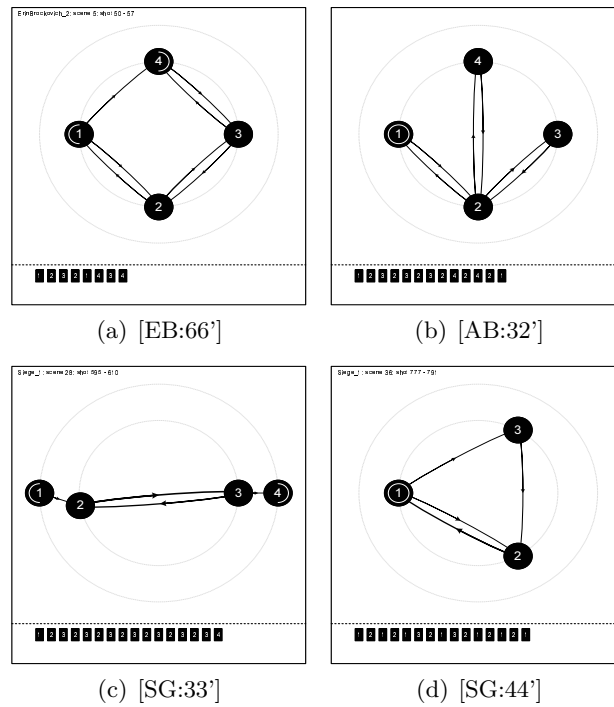


Figure 9: Editing orchestration examples.

Table 9: Verification of unit grouping and dialogue for examples in Fig. 9.

	[EB:66']	[AB:32']	[SG:33']	[SG:44']
U_l			1 \rightarrow 2	
U_r				
U_o			4 \rightarrow 3	
U_c				
GrpError				
Dlg.	(1,2),(2,3),(3,4)	(2,3),(2,4)	(2,3)	(1,2)
DlgError		(1,2)		

introductory long shot of the restaurant scene in [SG:33'] leads 9(d) logically to a closer shot of main character, Elise. The last \mathcal{O} -node maintains the movement continuity from the last shot of Take 3 in which the characters rise up from the chair.

In addition, most two-way connected \mathcal{I} -nodes in these examples depict the main dialogue in the scenes. Take pair (1,2) can be considered as an error because the second take is just a closer shot of the establishing shot in the first take. Dialogue lines are delivered in these takes but their two-way connection does not result from shot-reverse-shot editing.

5 Conclusions

In this paper, a visualization concept called the Double-Ring Take-Transition-Diagram is proposed for representing the internal structure of a scene. This technique presents takes and their transitions via nodes and edges of a ‘graph’ consisting of two rings as its backbone. The construction of the diagram, including node classification, connecting nodes via edges and unit linking, are motivated from the understanding of film grammar for shot arrangement. In addition, there are various semantic elements made explicit in this representation. The usefulness of DR-TTD should not only be understood from the content analysis and annotation perspective. DR-TTD can also aid in the authoring of film and video content. For example, an amateur filmmaker can use DR-TTDs to verify if the editing of a scene matches their dramatic/narrative intentions and adjust it accordingly.

As an alternative to this visualization approach, we are considering a full automatic approach in extracting semantics from editing patterns which takes into account the DR-TTD structural characteristics and other attributes such as camera movement, face existence, audio types (i.e., speech, music, silence, etc.).

References

- [1] C. Dorai and S. Venkatesh, “Computational Media Aesthetics: Finding meaning beautiful,” *IEEE Multimedia*, vol. 8, no. 4, pp. 10–12, October-December 2001.
- [2] B. T. Truong, C. Dorai, and S. Venkatesh, “Automatic scene extraction in motion pictures,” *IEEE Transactions in Circuits and Systems for Video Technology (CSVT)*, vol. 13, no. 1, pp. 5–15, Jan 2003.
- [3] D. Zhong, H. Zhang, and S.-F. Chang, “Clustering methods for video browsing and annotation,” in *Storage and Retrieval for Still Image and Video Databases IV*, 1996, pp. 239–246.
- [4] B. T. Truong, S. Venkatesh, and C. Dorai, “Application of computational media aesthetics methodology to extracting color semantics in film,” in *ACM Multimedia (ACMMM’02)*, France Les Pins, Oct 2002, pp. 339–342.
- [5] M. Yeung, B.-L. Yeo, and B. Liu, “Segmentation of video by clustering and graph analysis,” *Computer Vision and Image Understanding*, vol. 7, no. 1, pp. 94–109, July 1998.
- [6] Y. Rui, T. S. Huang, and M. S., “Constructing table-of-content for videos,” *ACM Multimedia System Journal: Special Issue in Multimedia Systems on Video Libraries*, vol. 7, no. 5, pp. 359–368, 1999.
- [7] E. Veneau, R. Ronfard, and P. Bouthemy, “From video shot clustering to sequence segmentation,” in *ICPR’00*, vol. 4, Barcelona, sep 2000, pp. 254–257.
- [8] Y. Matsuo, M. Amano, and K. Uehara, “Mining video editing rules in video streams,” in *ACMMM’02*, France Les Pins, Oct 2002.

-
- [9] M. Kumano, Y. Ariki, M. Amano, K. Uehara, K. Shunto, and K. Tsukada, "Video editing support system based on video grammar and content analysis," in *ICPR'02*, 2002, pp. 1031–1036.
 - [10] B. T. Truong, S. Venkatesh, and C. Dorai, "Identifying film takes for cinematic analysis," in *IEEE International Conference on Multimedia & Expo (ICME'03)*, vol. 2, Baltimore, 2003, pp. 405–408.
 - [11] B. T. Truong, "An investigation into structural and expressive elements in film," PhD Thesis, Department of Computing, Curtin University of Technology, Western Australia, apr 2004.