Video Content Analysis Tool:

3DVideoAnnotator

Graphical User Interface for annotating video sequences

Version 2.4

Table of Contents

1. Features 3
   1.1. Introduction 3
   1.2. Implementation 4
   1.3. Installing and Uninstalling 4
2. Graphical User Interface 4
   2.1. File Menu 4
   2.2. View Menu 5
   2.3. Windows Menu 7
      2.3.1. Player 7
      2.3.2. Annotator 9
         2.3.2.1. Shot Annotation 10
         2.3.2.2. Key Segment Annotation 11
         2.3.2.3. Event Annotation 12
         2.3.2.4. Object Annotation 12
         2.3.2.5. Human Annotation 14
      2.3.3. Timeline 17
      2.3.4. Editor 20
         2.3.4.1. Shot Editing 21
         2.3.4.2. Transition Editing 23
         2.3.4.3. Key Segment Editing 24
         2.3.4.4. Event Editing 26
         2.3.4.5. Static Object Editing 27
         2.3.4.6. Static Human Editing 29
         2.3.4.7. Moving Object Editing 31
         2.3.4.8. Moving Human Editing 33
         2.3.4.9. Cut Editing 36
         2.3.4.10. Header Editing 37
      2.3.5. Analyzer 38
         2.3.5.1. Shot Boundary Detector's Manual 40
         2.3.5.2. Haarcascade frontal face detector manual 40
         2.3.5.3. Color+Haarcascade Frontal Face Detector's Manual 41
         2.3.5.4. Frontal–Profile Face Detector's Manual 42
         2.3.5.5. Object Detector's Manual 43
         2.3.5.6. Particles Tracker's Manual 44
         2.3.5.7. LSK Stereo Tracker's Manual 45
         2.3.5.8. LSK Tracker's Manual 46
         2.3.5.9. 3D Rules Detector's Manual 47
         2.3.5.10. UFO Detector's Manual 48
         2.3.5.11. Keyframe Selection Tool's Manual 49

1. Features

1.1. Introduction

3DVideoAnnotator is an application that assists users in the task of annotating video sequences and viewing the corresponding results. A video sequence can be a single-view video or a stereoscopic video consisting of two channels (left and right) and their corresponding disparity channels. Each video can be annotated with shot descriptions, key segment descriptions, event descriptions, object and human (either static or moving) descriptions, either manually or automatic through algorithms. Users can navigate the descriptions through user friendly modules such as timelines and a tree view representation, and edit them. The application also allows the annotated descriptions to be stored in an output AVDP/XML file and can read existing descriptions from an AVDP/XML file. Figure 1 displays the application.

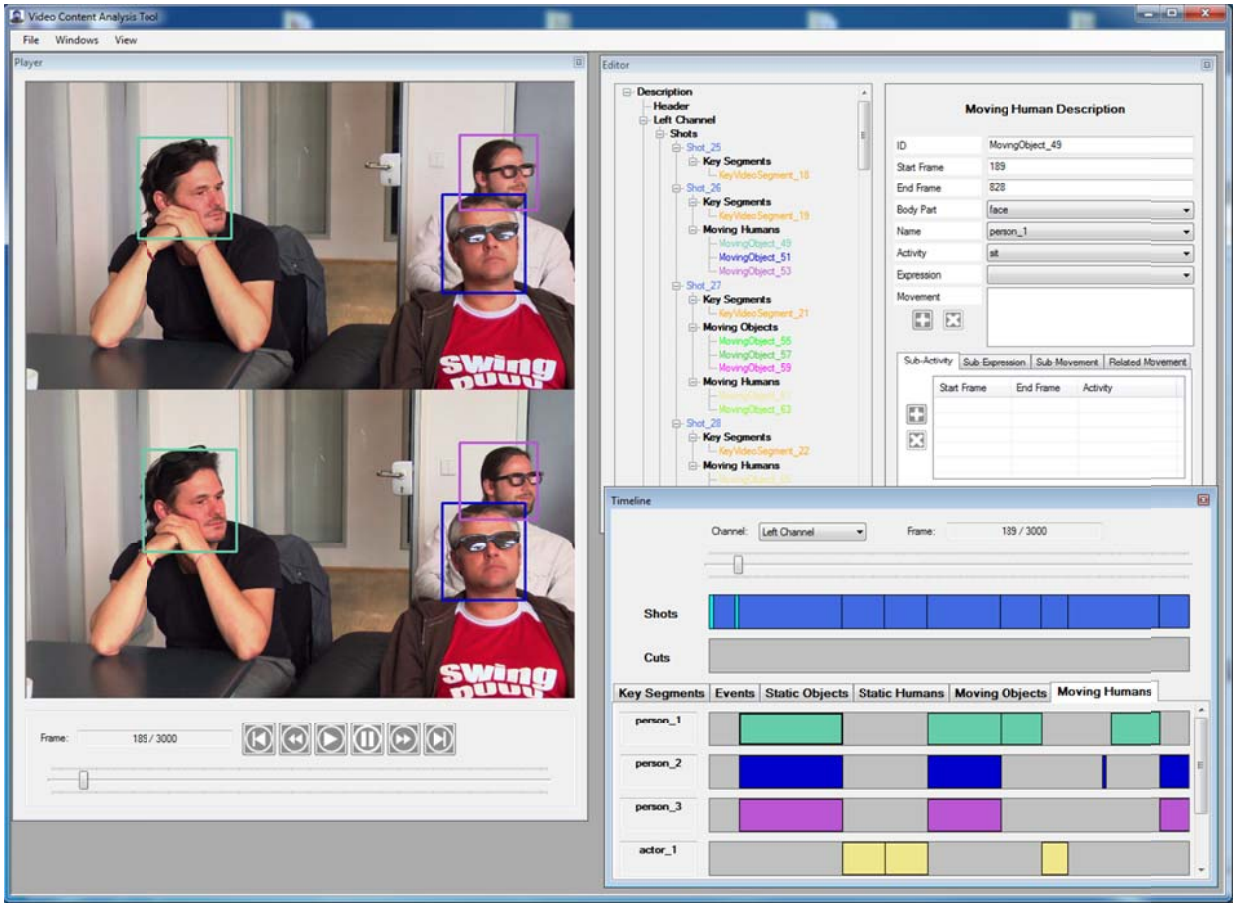


Figure 1: 3DVideoAnnotator application.

1.2. Implementation

3DVideoAnnotator is a Windows Forms Application. It is coded in C++/CLI programming language making use of the OPENCV library for handling videos and the XERCES library for parsing the XML files. Libraries, which implement various video content analysis algorithms and the storage of descriptions in AVDP/XML files, are used as well. It is a multiple-document interface (MDI) application, where all main operations (e.g., manual annotation, navigation of video and audio content's descriptions) are executed through separate forms. GUI components such as buttons, sliders and drop-down menus are used in order to provide user friendliness and ease of use.

1.3. Installing and Uninstalling

This is a stand-alone application, which means no installation is required. All application's required files reside inside the root folder which can be extracted anywhere in the user's computer. The program requires .NET Framework 4 and Microsoft Visual C++ 2010 Redistributable Package to be installed on the computer.

Uninstalling the application can be performed by simply erasing the root folder.

2. Graphical User Interface

All operations are executed through menus, which are described below.

2.1. File Menu

Through the *File Menu*, the user can open an AVI file, save/load a video content's description, etc.. The menu contains the following functions, as shown in Figure 2.

- **Open Single Video** – It opens an AVI file.
- **Open Stereo Video → Two videos (L/R)** – It opens two dialog boxes though the user selects the left and the right channel of a stereoscopic video, respectively.
- **Open Stereo Video → Four videos (L/R plus Disparity)** – It opens four dialog boxes though the user selects the left and the right channel of a stereoscopic video and the respective disparity channels, respectively.
- **Open Stereo Video → One video → Left-Right** – It opens an AVI file corresponding to a stereoscopic video which contains the two channels side-by-side.

- **Open Stereo Video → One video → Top-Bottom** – It opens an AVI file corresponding to a stereoscopic video which contains the two channels in a top-bottom manner.
- **Open Stereo Video → One video → Left-Right plus Disparity** – It opens an AVI file corresponding to a stereoscopic video which contains the color and disparity channels side-by-side.
- **Open Stereo Video → One video → Top-Bottom plus Disparity** – It opens an AVI file corresponding to a stereoscopic video which contains the color and disparity two channels in a top-bottom manner.
- **Read XML (AVDP)** – It loads a video content's description from an AVDP/XML file. The loaded description is added to the existing description.
- **Save XML (AVDP)** – It saves the video annotations as an AVDP/XML file.
- **Exit**

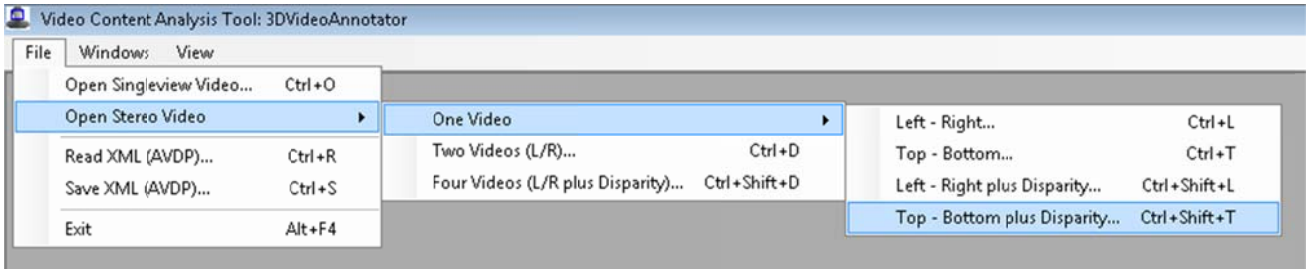


Figure 2: File menu.

2.2. View Menu

The *View Menu* (Figure 3) is associated with the video viewing. Specifically:

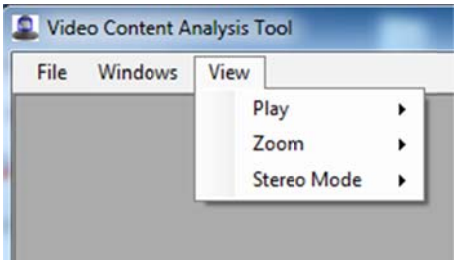


Figure 3: View Menu.

- **Play** – It includes the following three playback modes (Figure 4):
  - **Slow –** Play the video in slow mode.
  - **Normal –** Play the video according to its frame rate.
  - **Fast –** Play the video in fast mode.

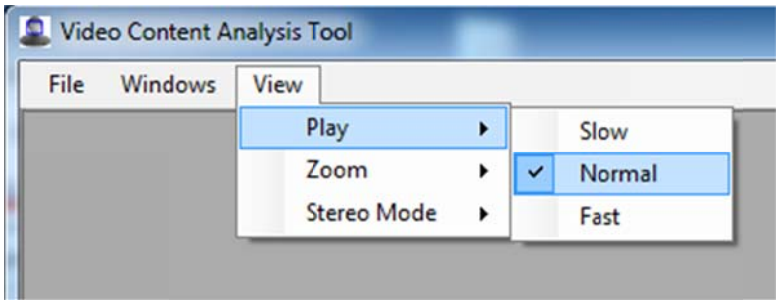


Figure 4: Playback modes.

- **Zoom** – It includes seven available zoom factors as shown in Figure 5. Note that it is possible for some zoom factors to be disabled, if the screen resolution is low or if the frame size is large.

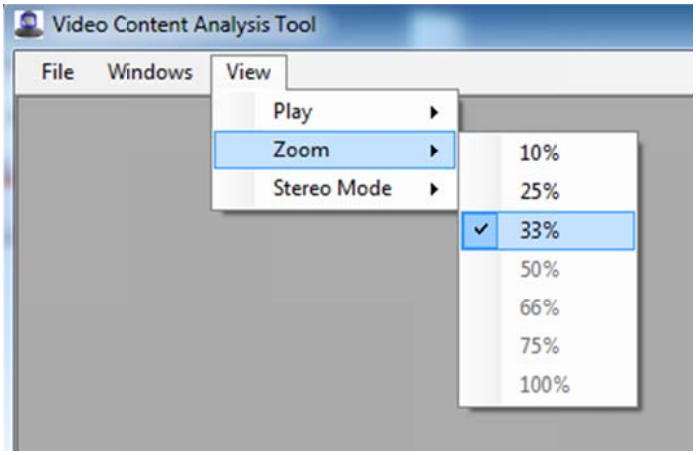


Figure 5: Zooming.

- **Stereo Mode** – If the input video is a stereoscopic video, the user can select which of the channels will be visible in the *Player Window* through the modes shown in Figure 6.

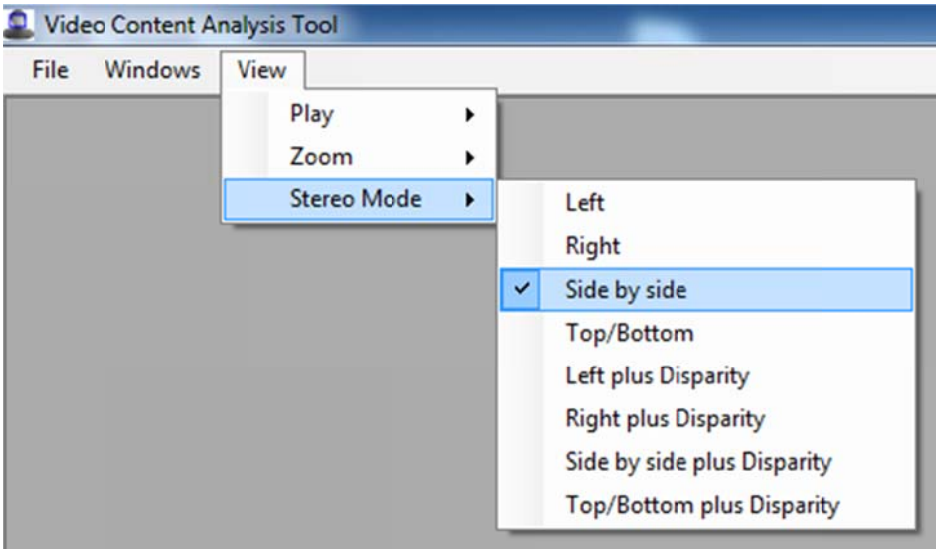


Figure 6: Stereo Modes.

2.3. Windows Menu

The *Windows Menu* (Figure 7) is the most important one since all main operations (video playback, video content's description annotation, editing and navigation) are initiated here.

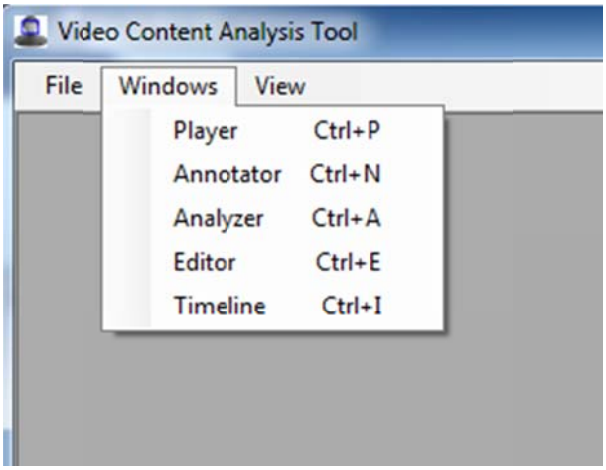


Figure 7: Windows Menu.

Each of the five windows is described next.

2.3.1. Player

The *Player Window* opens a video player with navigation buttons and slider, as depicted in Figure 8:

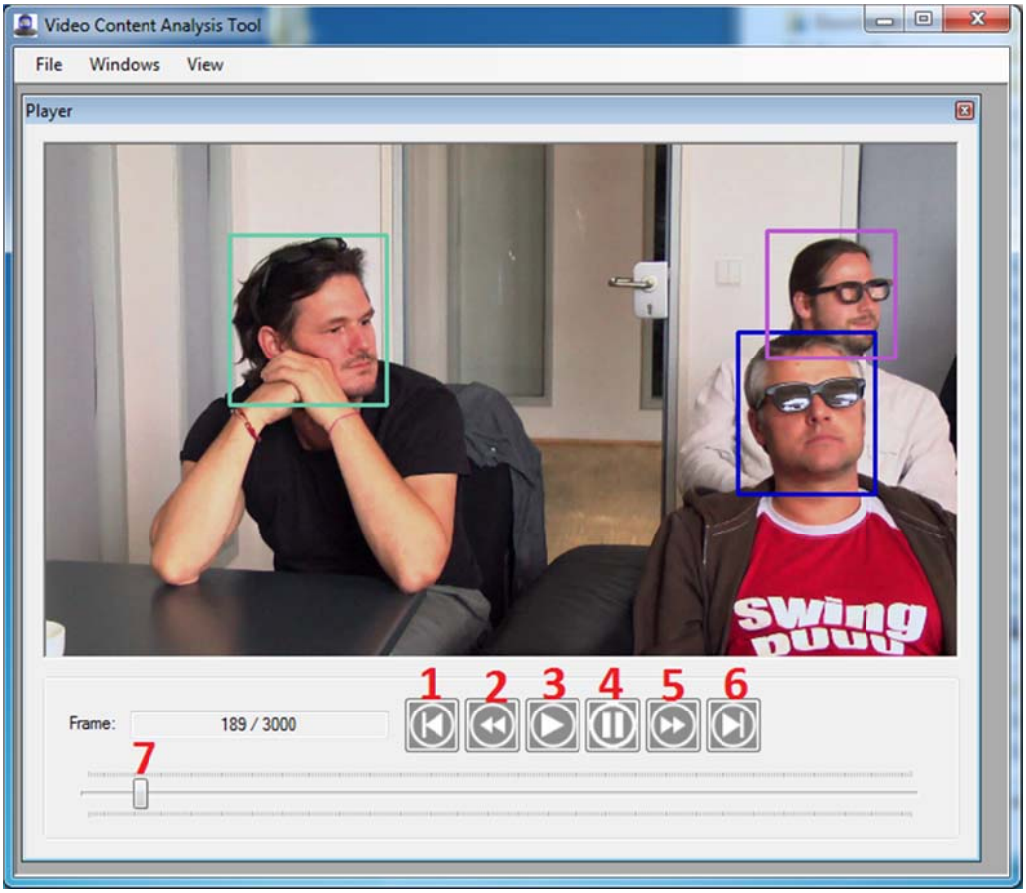


Figure 8: Player Window.

The functionality of the buttons of the video player (Figure 8) is explained below:

1. It moves to the start of the video.
2. It moves to the first frame of the previous shot.
3. It starts playback of the video.
4. It stops playback.
5. It moves to the first frame of the next shot.
6. It moves to the end of the video.
7. By dragging the slider the user can navigate through the video.

Note that the first frame is frame number 1.

The video player also handles some mouse events. If the user presses the right button of the mouse at any position on the video player, a dropdown menu will appear through which the video player can be resized (Figure 9).

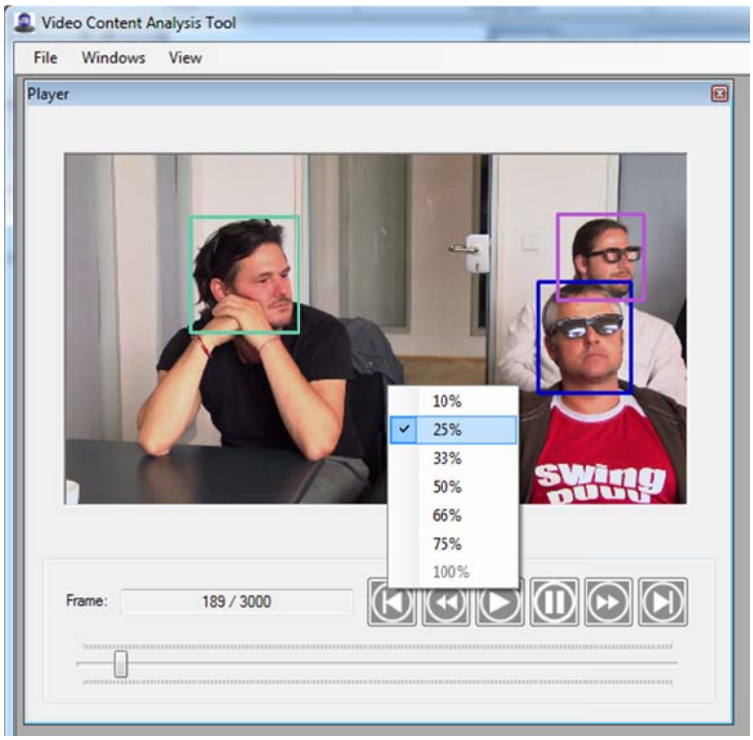


Figure 9: Zooming in the video player.

If the user double-clicks within a bounding box, then the description of the corresponding moving/static object (or human) is presented on the *Editor Window*. If the user double-clicks within the frame but outside the bounding boxes, the description of the corresponding shot (or transition) appears on the *Editor Window*. For more details see the section *2.3.4 Editor*.

Finally, the bounding boxes, which are displayed on the video player, can be moved or resized by mouse-click events, as depicted in Figure 10.

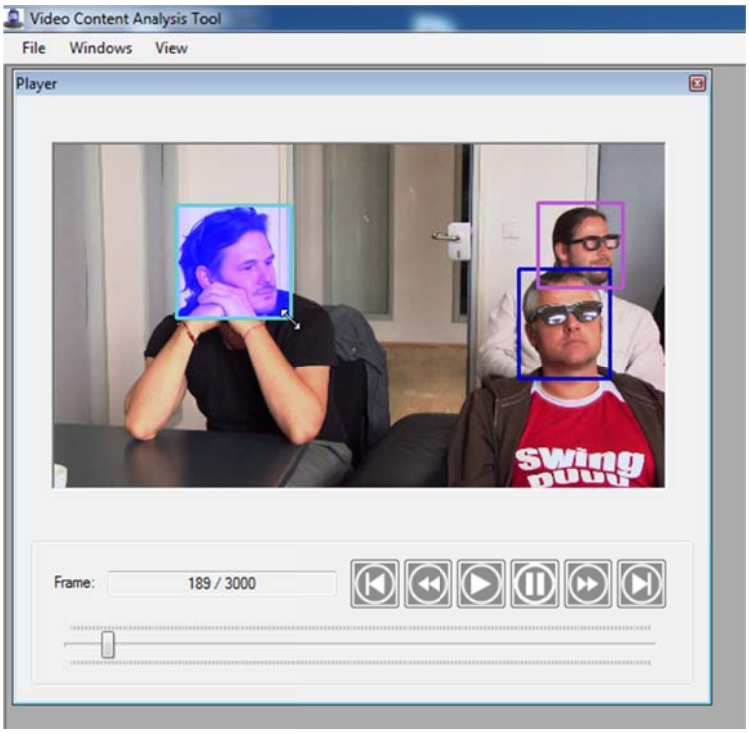


Figure 10: Bounding box editing.

2.3.2. Annotator

The *Annotator Window* (Figure 11) gives the ability to manually annotate the video content.

Figure 11: Annotator Window.

Specifically, each video can be annotated with descriptions of shots, key segments, events, objects (moving and static) or humans (moving and static). First the user selects the type of annotation he/she wishes to perform (see Figure 11) and then proceeds with the annotation. Additionally, the user can define on which channels the annotation will be applied. Each of the annotation types is presented in detail next.

2.3.2.1. Shot Annotation

Pressing the *Shot button*, the user can start annotating a shot.

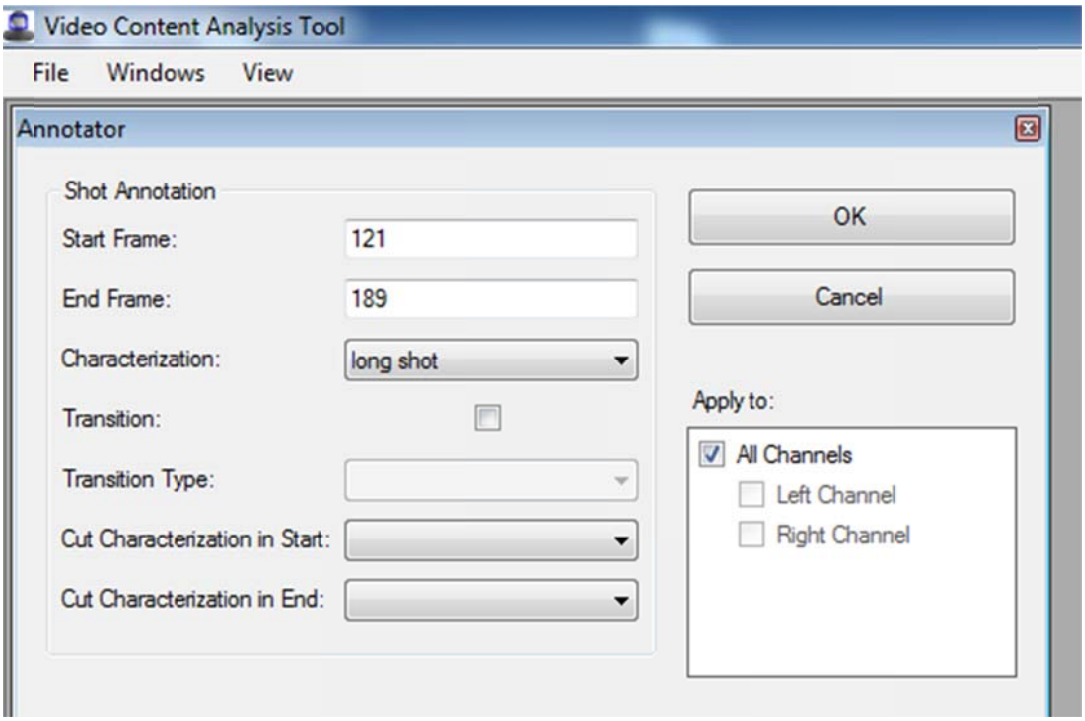


Figure 12: Shot Annotation.

According to Figure 12, the user is able to define the following attributes:

- **Start Frame** - The first frame of the shot. It is initialized to the frame number where the annotation started.
- **End Frame** - The last frame of the shot. It is updated to the frame number of the current frame shown.
- **Characterization** - The shot can be characterized with terms, such as close-up or comfortable for viewing, by selecting a characterization from the corresponding drop-down list.
- **Transition** - If the user wants to annotate a series of frames as being a transition, the corresponding checkbox should be checked.

- **Transition Type** - The type of the transition (such as cross-dissolve or fade-in) should be selected from the corresponding dropdown list, if a transition is being annotated.
- **Cut Characterization in Start** - The start of the shot is automatically annotated as a cut. Optionally, the cut can be characterized with characterizations such as comfortable or uncomfortable, using the corresponding drop-down list.
- **Cut Characterization in End** - The end of the shot is automatically annotated as a cut. Optionally, the cut can be characterized with characterizations such as comfortable or uncomfortable, using the corresponding drop-down list.

Note that all drop-down lists contain some pre-defined terms. However, it is possible for the user to define and add new terms. In order to ensure that the entire video will consist of no-overlapping shots (or transitions), any new shot annotation causes changes to the duration of the existing shots. For example, if a video with 100 frames consists of two shots (the first one starts from frame 1 and has duration 45 frames, while the second one starts from frame 46 and ends in frame 100) and a new shot (from frame 30 to frame 70) is inserted, the video will consist of the following shots: the first shot will start from frame 1 and end in frame 29, the second one will start from frame 30 and end in frame 70 and the third one will start from frame 71 and end in frame 100.

2.3.2.2. Key Segment Annotation

In order to annotate a frame or a series of frames, as being a key frame or a key video segment respectively, the user should press the *Key Segment button* and simply set its duration, as depicted in the Figure 13.

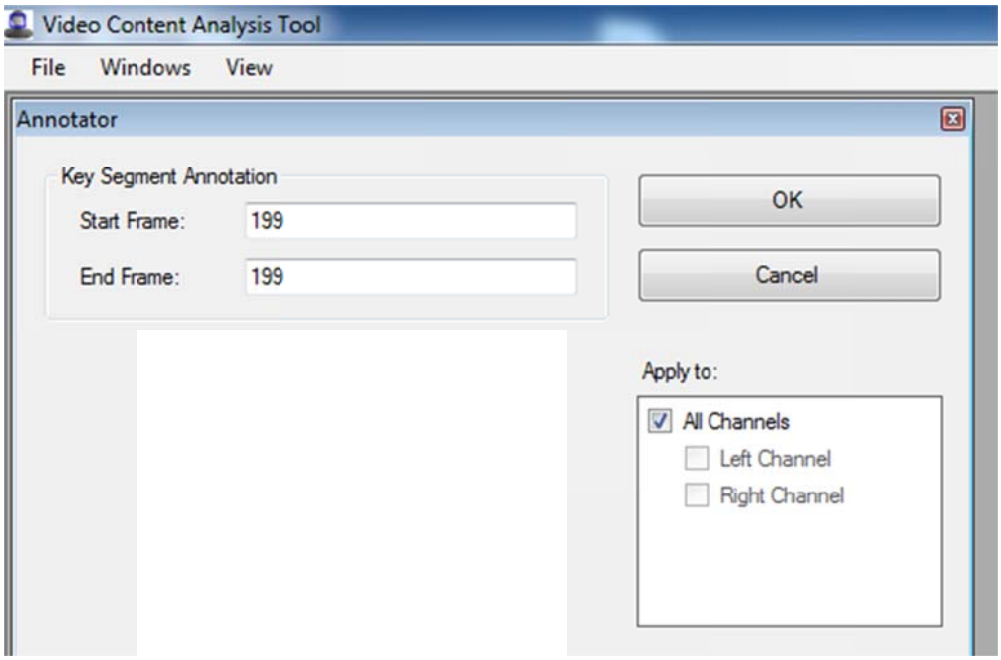


Figure 13: Key Segment Annotation.

2.3.2.3. Event Annotation

Pressing the *Event button*, the user can start annotating an event.

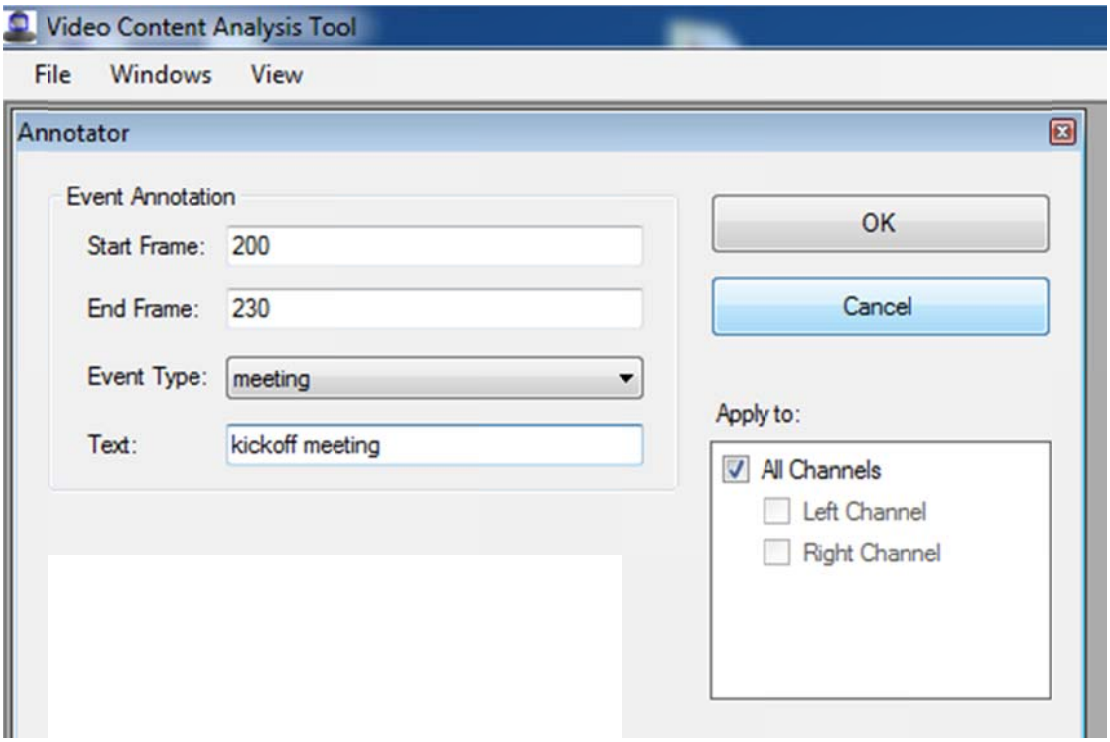


Figure14: Event Annotation.

According to Figure 14, the user is able to define the following attributes:

- **Start Frame** - The first frame of the event. It is initialized to the frame number when the annotation started.
- **End Frame** - The last frame of the event. It is updated to the frame number of the current frame shown.
- **Event Type** - The type of the event should be defined either by selecting a pre-defined term from the drop-down list or by adding a new one.
- **Text** – A free text can be added, so as to describe the event by using natural language.

2.3.2.4. Object Annotation

Pressing the *Object button*, the user can start annotating an object. By object, we mean any region on a frame which does not correspond to a person. Firstly, the user should define the object appearance over one or more frames. The region of a static object on a frame is defined by clicking-and-dragging the mouse to create a bounding box over a video frame. If the user wants to annotate a moving object, he/she must draw the bounding boxes in subsequent frames. The application generates the bounding boxes in intermediate frames of the same video channel (e.g. the left one) and in the same

positions in the other video channels (e.g. the right one). Then, the bounding boxes, which are displayed on the video, can be moved or resized by mouse-click events.

As seen in Figure 15, the user is able to define the following attributes for a static object:

- **Object Class** - The object class, such as chair or car, is defined by selecting a term from the corresponding drop-down list.
- **Object Type** - The specific type of the object, for example an office chair.
- **Orientation** - The orientation of the object, such as oriented left.
- **Position -** The position of the object, such as left.
- **Size -** The size of the object, such as small.

Note that the four last attributes are optional. Also, all drop-down lists contain some pre-defined terms, while it is possible for the user to define and add new terms.

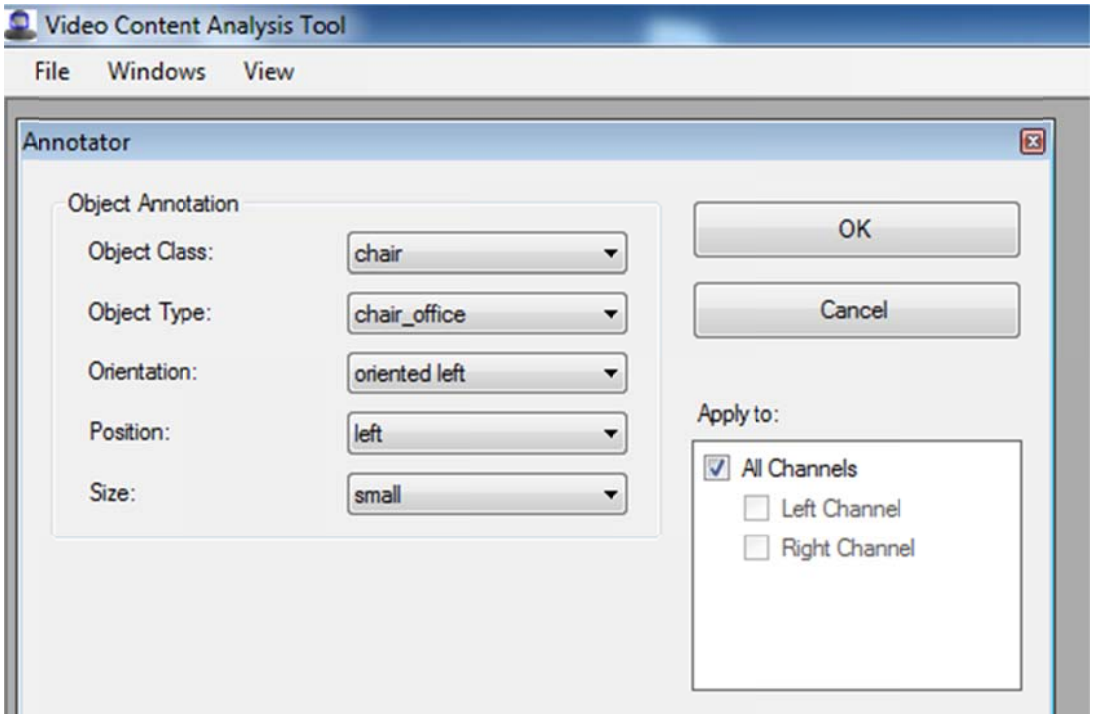


Figure 15: Static Object Annotation.

If a moving object is being annotated, according to Figure 16, the user is able to define the following attributes:

- **Object Class** - The object class, such as ball, is defined by selecting a term from the corresponding drop-down list.
- **Object Type** - The specific type of the object, for example a football ball.
- **Movement Direction** - The movement direction of the object, such as left.
- **Movement Speed -** The movement speed of the object, such as fast.

- **Sub-Movement -** In case different movements occur within the same object appearance, i.e. within a consecutive set of frames where the object appears, the user can define their durations and specific characteristics.

Note that only the Object Class is obligatory to be set.

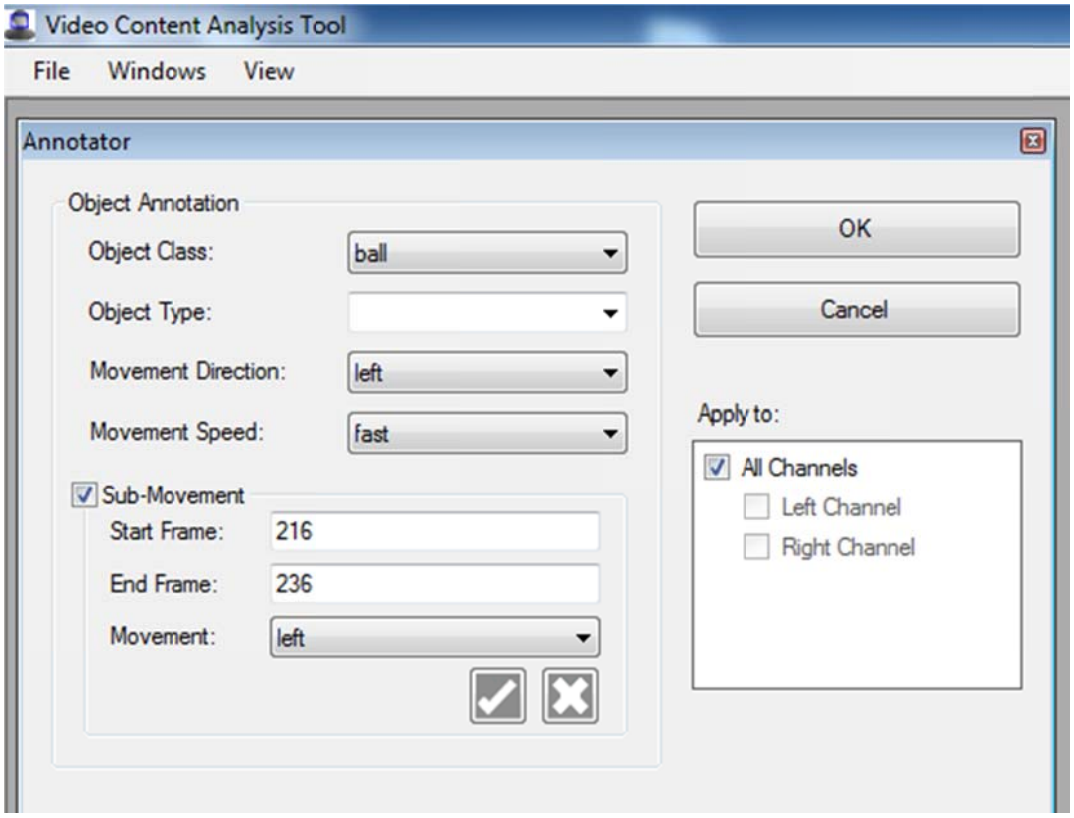


Figure 16: Moving Object Annotation.

2.3.2.5. Human Annotation

By pressing the *Human button*, the user can start annotating a human, i.e. any region on a frame which corresponds to a person, such as body, face. Firstly, the user should define the human appearance over one or more frames. The region of a static human on a frame is defined by clicking-and-dragging the mouse to create a bounding box over a video frame. If the user wants to annotate a moving human, he/she should must the bounding boxes in subsequent frames. The application generates the bounding boxes in intermediate frames of the same video channel (e.g. the left one) and in the same positions in the other video channels (e.g. the right one). Then, the bounding boxes, which are displayed on the video, can be moved or resized by mouse-click events.

As seen in Figure 17, the user is able to define the following attributes for a static human:

- **Body Part** - The body part of the human actor enclosed in the bounding box is defined by selecting a term from the corresponding drop-down list.

- **Name** - The name of the human. This can refer to either an actual name (e.g., Bogart) or a symbolic name (e.g., person_1).
- **Activity** - The activity of the human, such as walk.
- **Expression** - The facial expression of the human, such as anger.
- **Orientation** - The orientation of the human, such as oriented left.
- **Position -** The position of the human, such as left.
- **Size -** The size of the human, such as small.

Note that the six last attributes are optional. All drop-down lists contain some pre-defined terms, while it is possible for the user to define and add new terms.

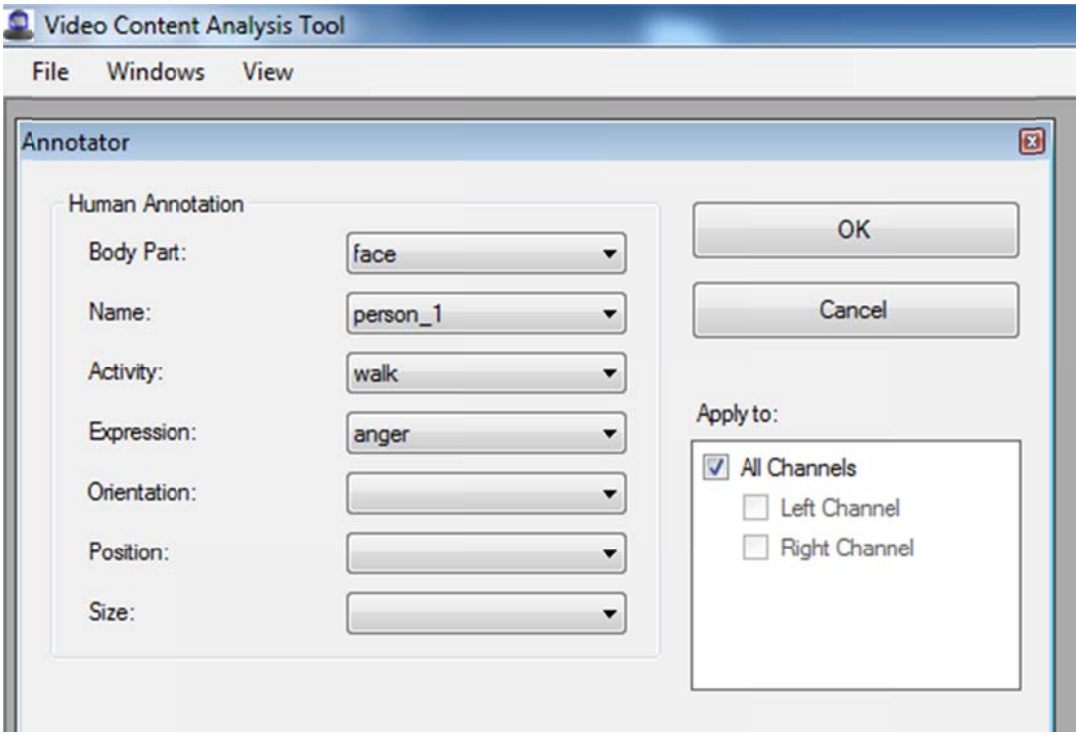


Figure 17: Static Human Annotation.

If a moving human is being annotated, then according to Figure 18 the user can define the following attributes:

- **Body Part** - The body part of the human actor enclosed in the bounding box is defined by selecting a term from the corresponding drop-down list.
- **Name** - The name of the human. This can refer to either an actual name (e.g., Bogart) or a symbolic name (e.g., person_1).
- **Activity** - The activity of the human, such as walk.
- **Expression** – The facial expression of the human, such as anger.
- **Movement Direction** - The movement direction of the human, such as left.

- **Movement Speed -** The movement speed of the human, such as fast.
- **Sub-Activity -** In case different activities occur within the same human appearance, the user can define their durations and specific activities.
- **Sub-Expression -** In case different expressions occur within the same human appearance, the user can define their durations and specific expressions.
- **Sub-Movement -** In case different movements occur within the same human appearance, the user can define their durations and specific movements.

Note that only the Body Part is obligatory to be set.

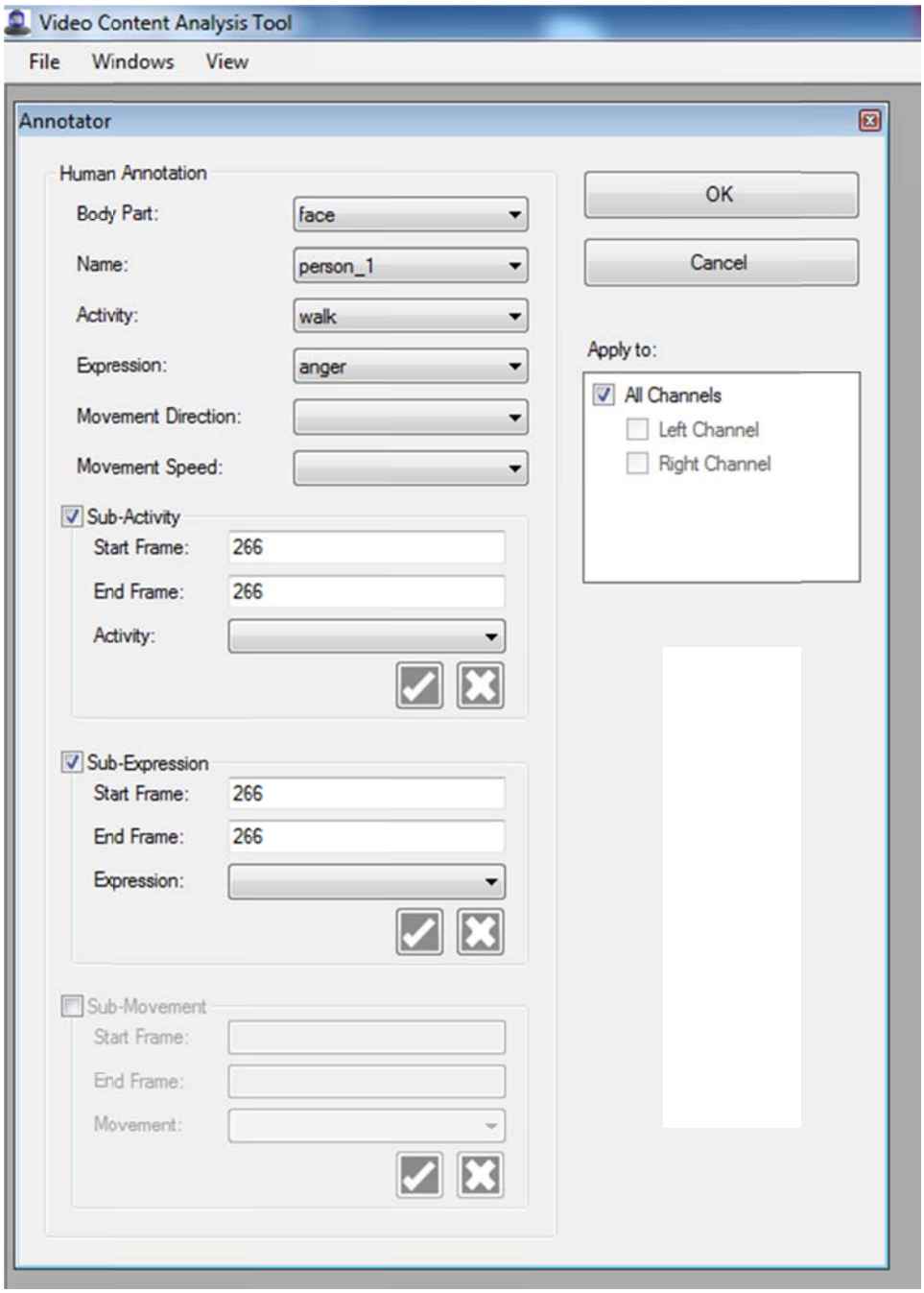


Figure 18: Moving Human Annotation.

2.3.3. Timeline

The *Timeline Window* (Figure 19) provides a user friendly way to view time-related parts of the video and audio content's description. Specifically, the user can navigate the descriptions of the shots, transitions and cuts (Figure 19), key frames and key video segments (Figure 19), events (Figure 20), static objects (Figure 21) and humans (Figure 22), moving objects (Figure 23) and humans (Figure 24) and audio sources. The descriptions are represented by colored areas on the Timeline. The length of each area shows the duration of the corresponding shot, event, etc.

The descriptions are organized based on semantic information. Thus, the events are grouped based on the type of the event (Figure 20), while the static and moving objects (or humans) are grouped based on their object type and name respectively (Figure 21-24), if this information is available. Otherwise, the unique ID is used. Finally, the audio sources are grouped based on the type of the source.

The user can navigate the descriptions per channel by selecting the channel he/she wishes to inspect from the corresponding drop-down list. When the mouse hovers over an area of the Timeline (e.g. on a shot or a moving period appearance), the id of the corresponding entity appears (Figure 19). Additionally, if the user clicks on an area of the Timeline, the corresponding description appears on the *Editor Window* (see the section *2.3.4 Editor*), while the first frame of the description is displayed on the video player. Also, in the case of static or moving objects and humans, the overlapping bounding box on the frame is shaded. Finally, the *Timeline Window* can be resized, in order for timelines to be shown in more detail.

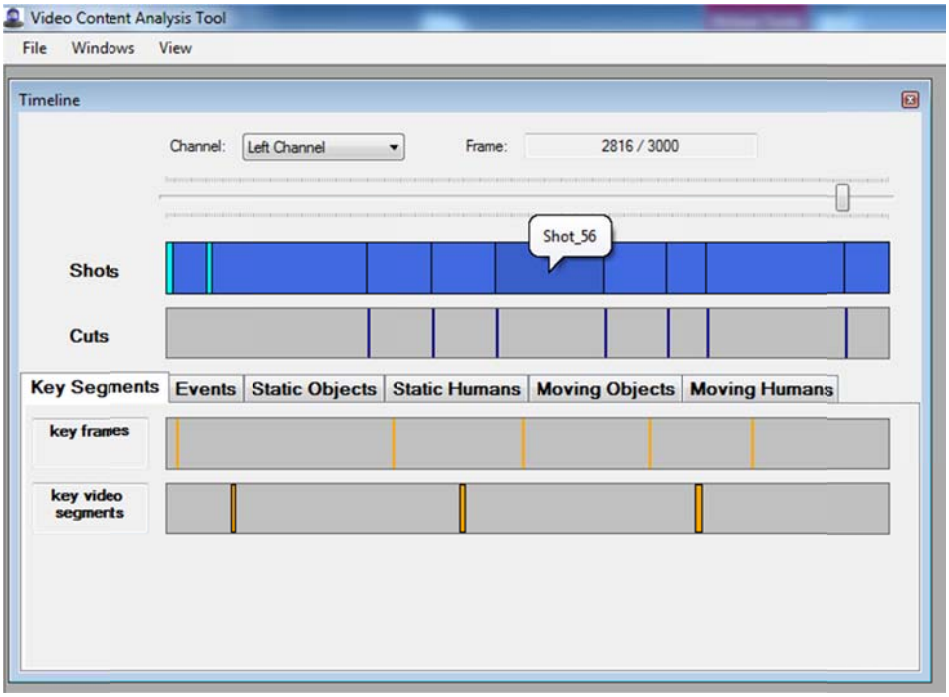


Figure 19: Shots and Key Segments on the Timeline Window.

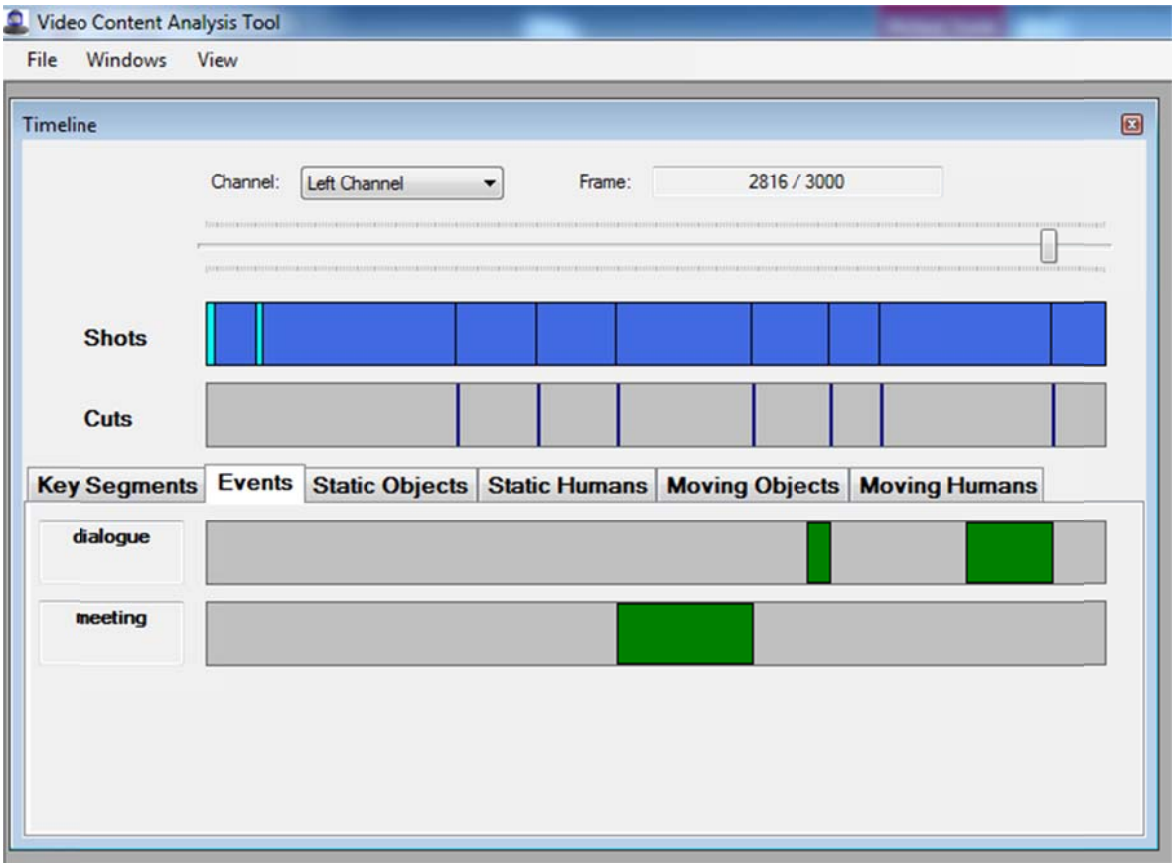


Figure 20: Events on the Timeline Window.

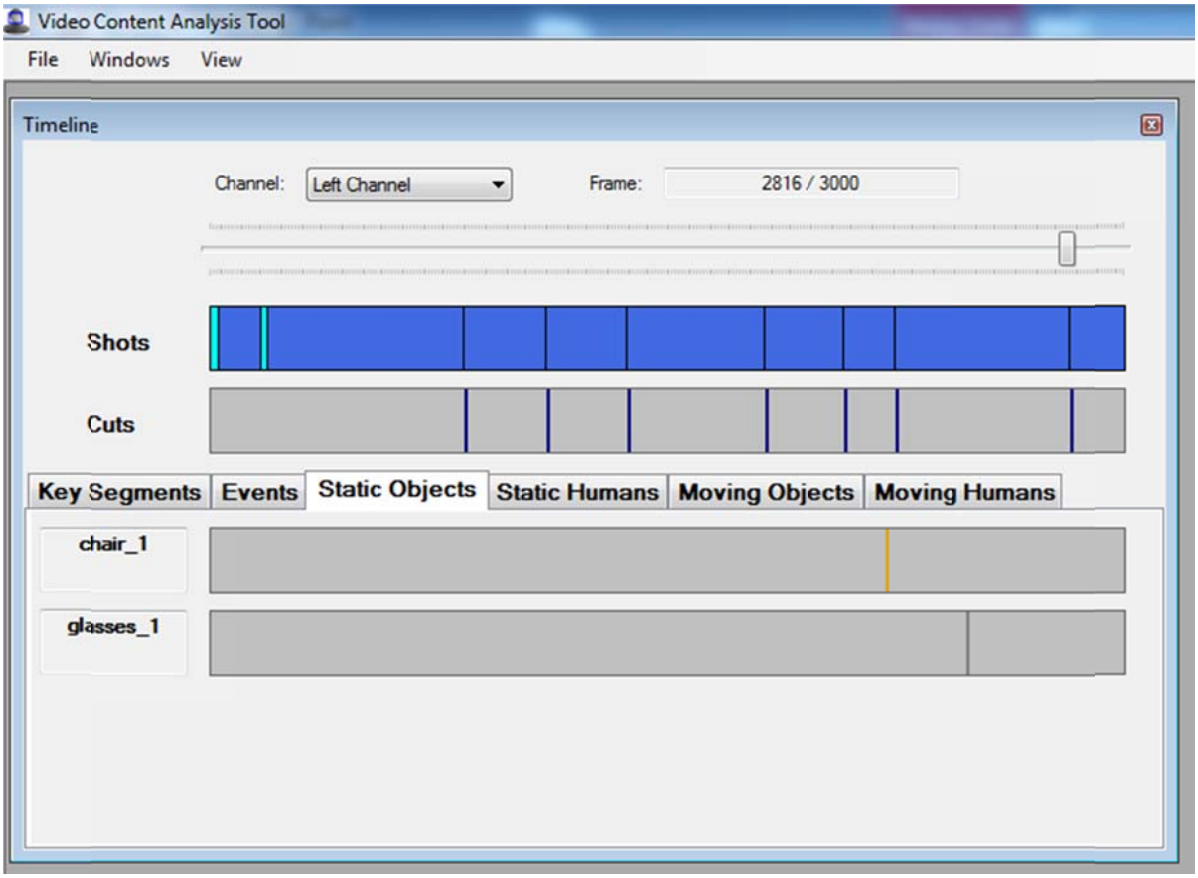


Figure 21: Static Objects on the Timeline Window.

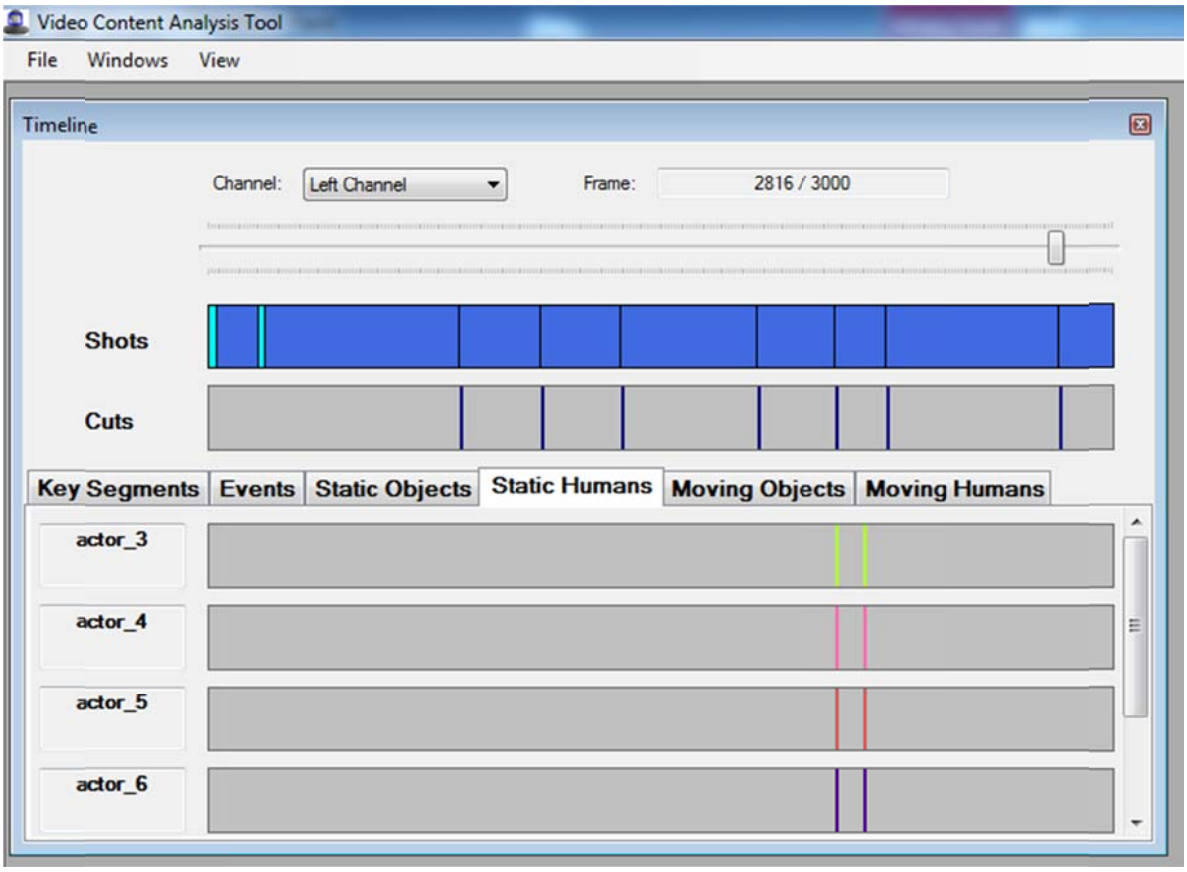


Figure 22: Static Humans on the Timeline Window.

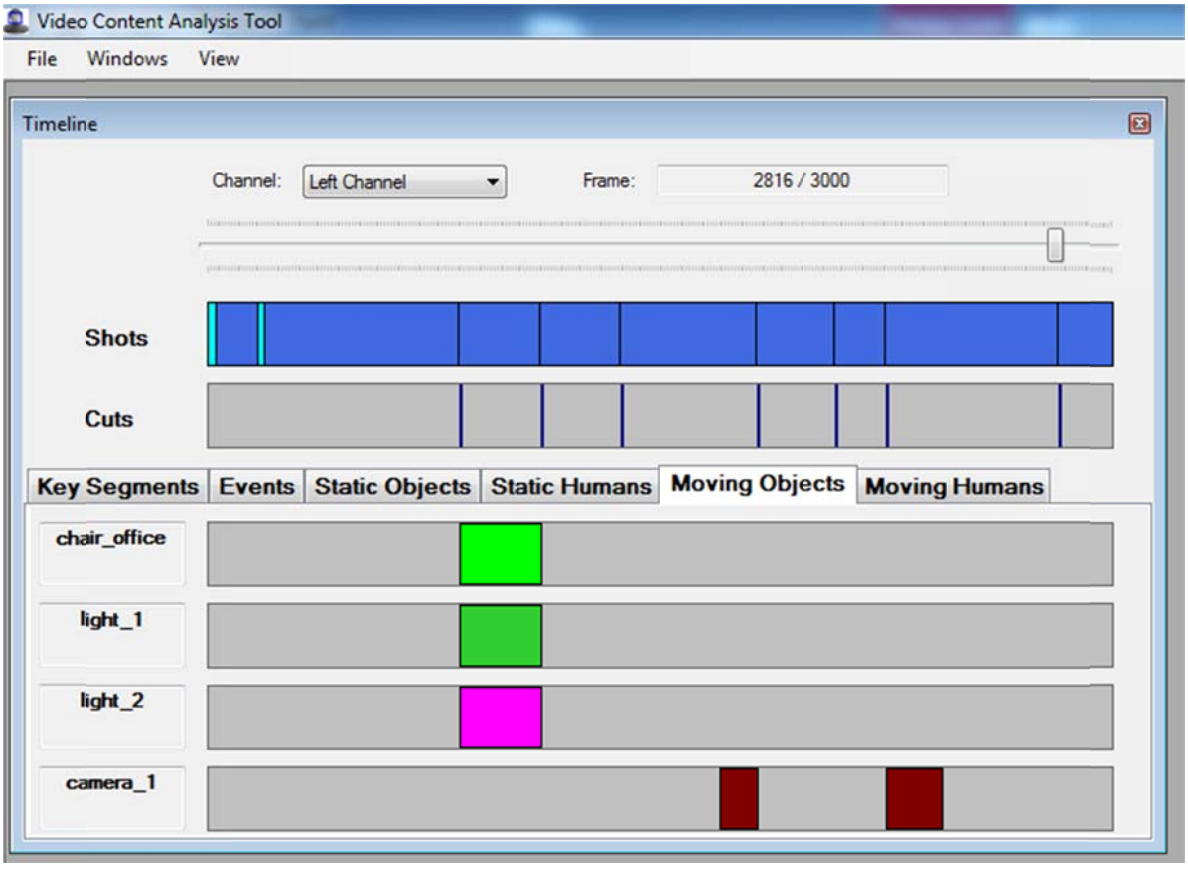


Figure 23: Moving Objects on the Timeline Window.

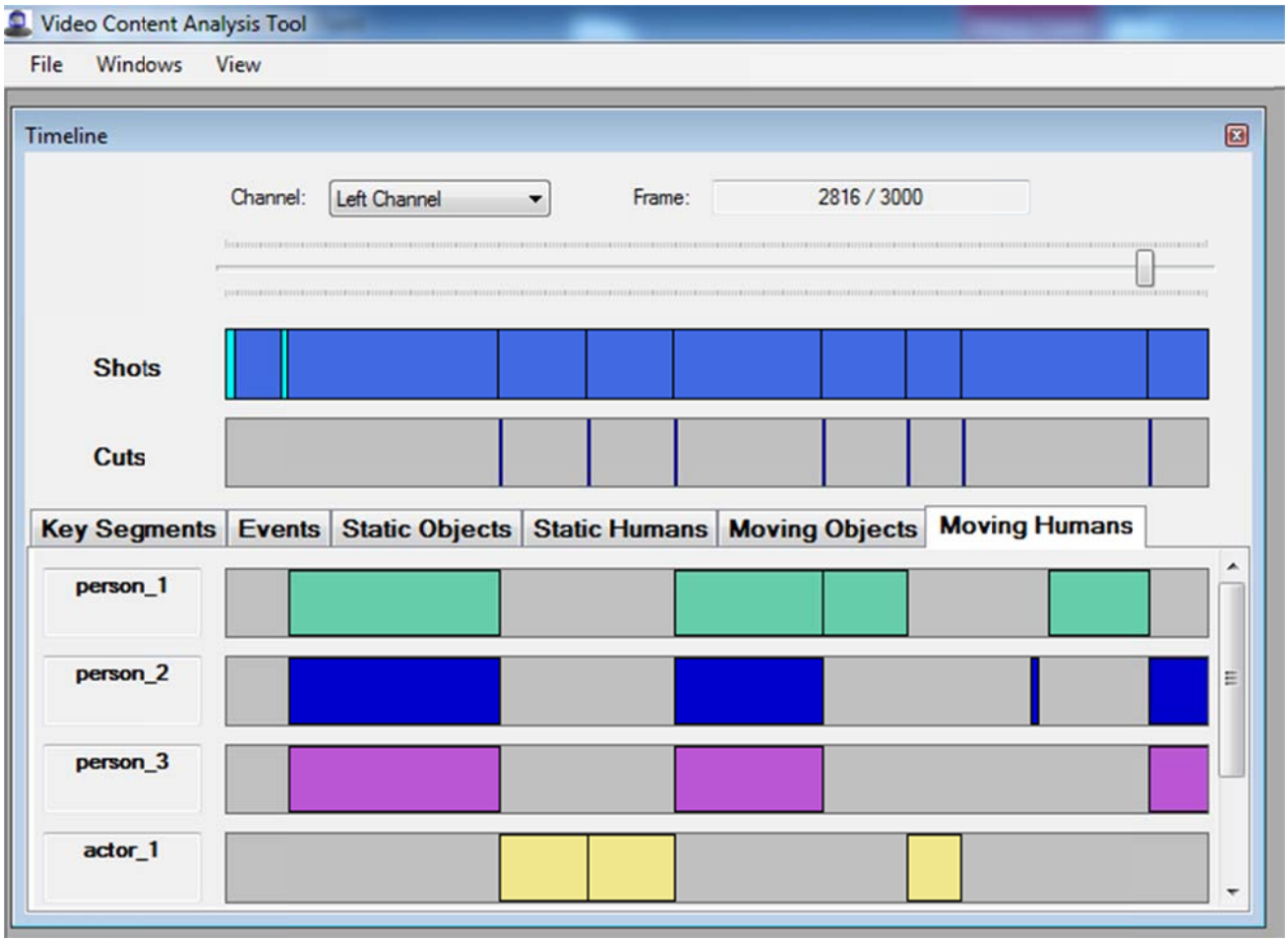


Figure 24: Moving Humans on the Timeline Window.

2.3.4. Editor

The *Editor Window* (Figure 25) provides an alternative way to navigate and edit the video content's description. The left part of the window displays the description in a hierarchical tree view form, making it structured and easy to use. The description is initially divided into header and channels, while each channel consists of two groups which contain the shots, transitions and cuts, respectively. Each of them contains in groups the key segments (key frames and key video segments), the events, the static and moving objects and humans, as depicted in Figure 25. Also, if audio description is available, an extra node is added to the left part, which contains nodes for viewing the audio sources. Each part of the description can be expanded or collapsed. When these nodes are double-clicked, the first frame of the corresponding description is displayed on the video player, while in the case of the static and moving objects and humans, the corresponding bounding box is shaded. Also, if their appearance spans more than one shots (or transitions) they are not saved in an AVDP/XML file and the node's label is underlined. Through the right part of the window the user can see and edit the description for each element of the above groups. Each group is presented in detail next.

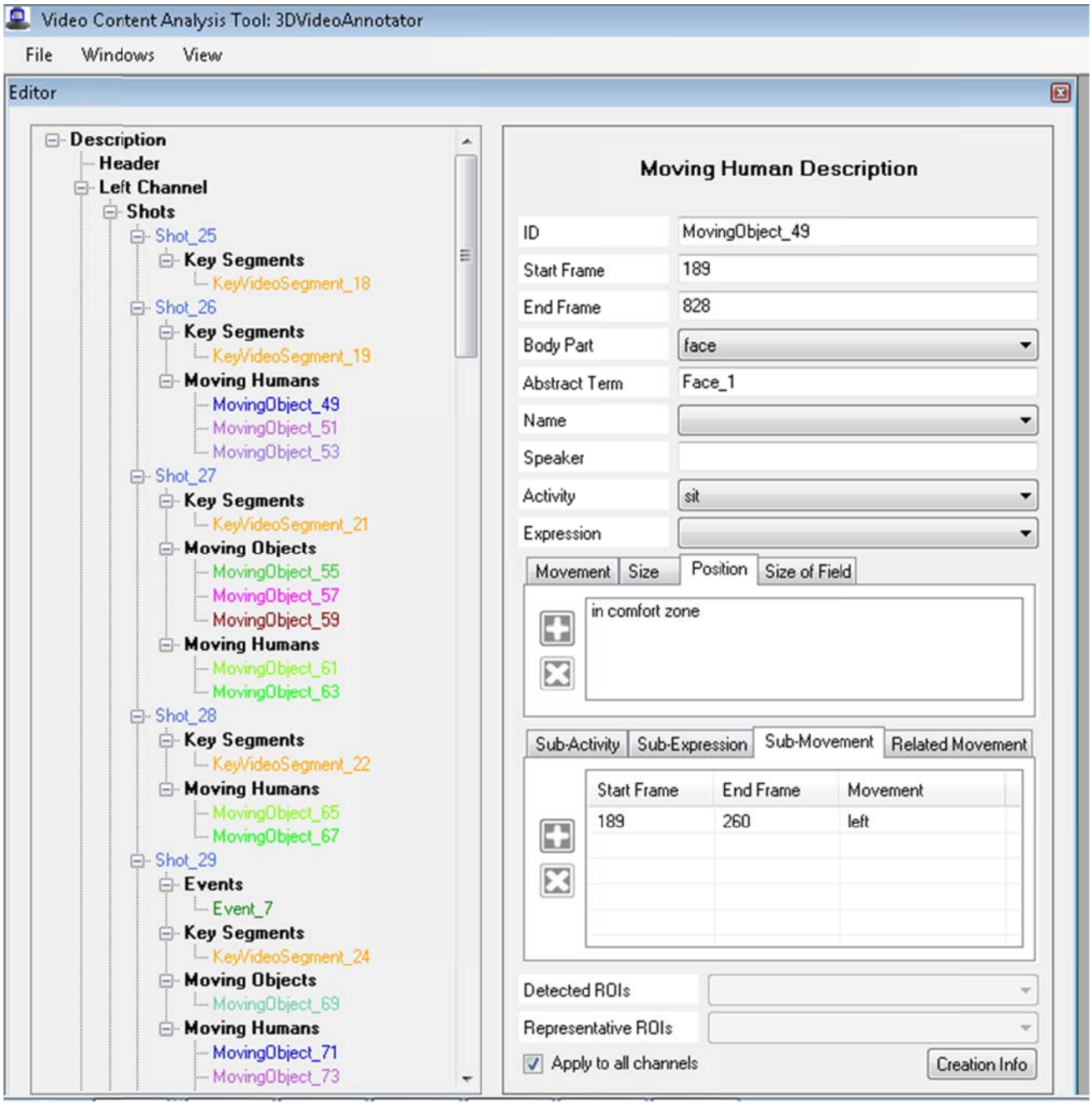


Figure 25: Editor Window.

2.3.4.1. Shot Editing

By left clicking on a node which represents a shot, the right part of the window displays the shot's description. So, according to Figure 26, the user can see and edit the following attributes:

- **ID** - The unique id of the shot. The value cannot be changed.
- **Start Frame** - The first frame of the shot.
- **End Frame** - The last frame of the shot.

- **Characterization** - The shot can be characterized with terms, such as close-up or comfortable for viewing, by selecting a characterization from the corresponding drop-down list. New terms can be added.
- **Spatial Spread of Objects -** The spread of many static objects on a frame, that is characterized with terms as spread or concentrated. Adding and deleting description terms is possible through the corresponding buttons. The changes can be applied to all the channels in the respective frames. Also, for each term a confidence level and a text to save extra information about the term can be stored by double-clicking on the term, as shown in Figure 27.

Note that any change to the start and/or the end frame of the shot causes changes to the duration of the other shots/transitions, as described in Section *2.3.2.1. Shot Annotation.*

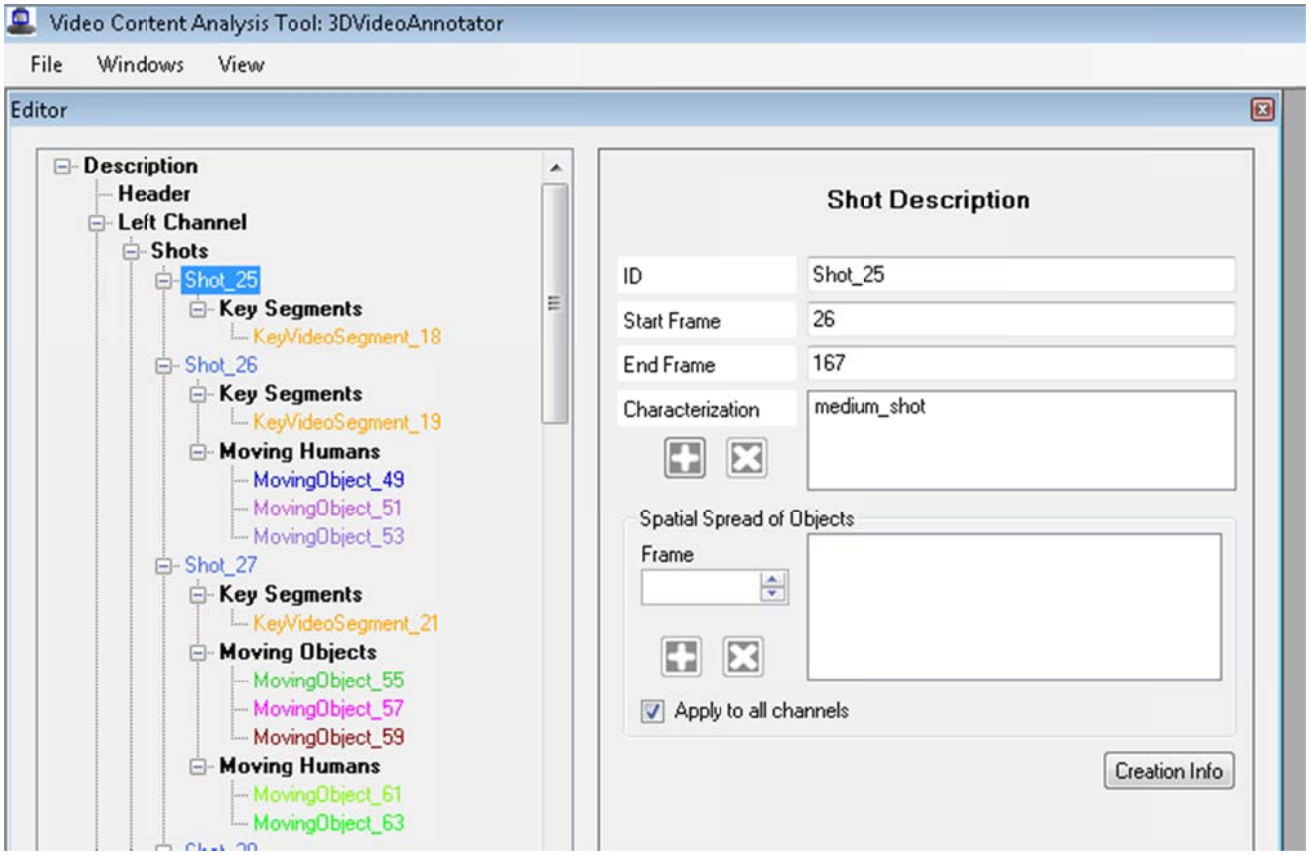


Figure 26: Shot Editing.

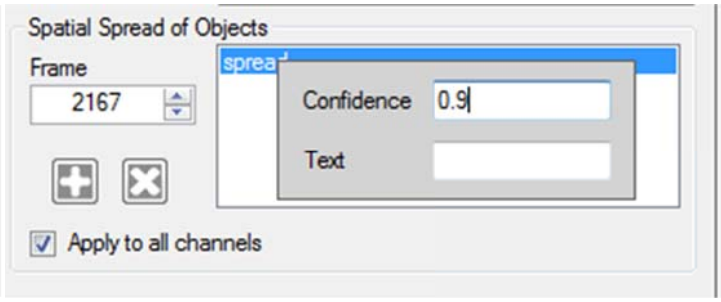


Figure 27: Spread Editing.

2.3.4.2. Transition Editing

By left clicking on a node which represents a transition, the right part of the window displays the transition's description.

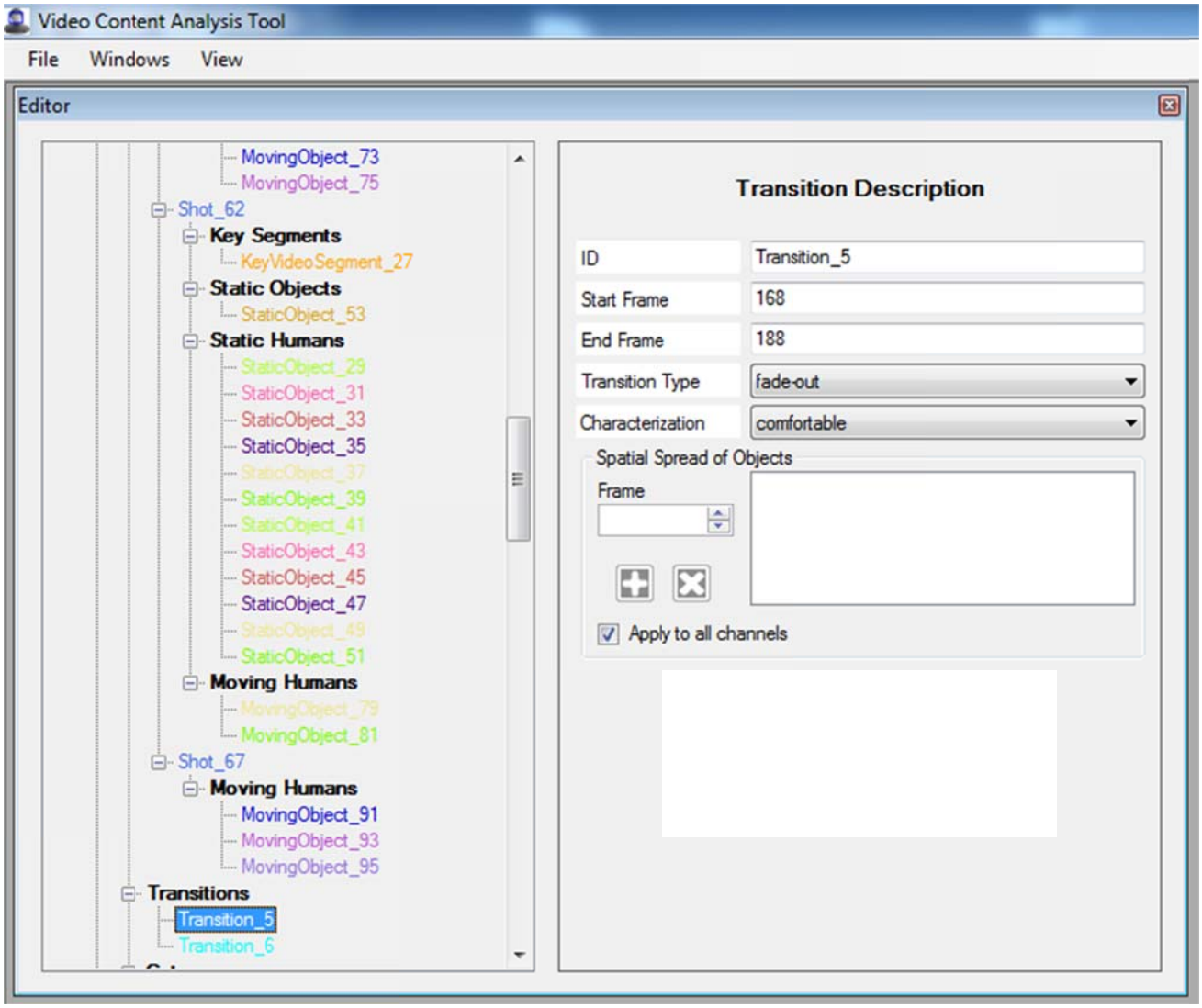


Figure 28: Transition Editing.

According to Figure 28, the user can see and edit the following attributes:

- **ID** - The unique id of the transition. The value cannot be changed.

- **Start Frame** - The first frame of the transition.
- **End Frame** - The last frame of the transition.
- **Transition Type** - The type of the transition (such as cross-dissolve or fade-in). The value can be changed by selecting a new term from the corresponding drop-down list. New terms can be added.
- **Characterization** - The transition can be characterized with terms, such as comfortable for viewing, by selecting a characterization from the corresponding drop-down list. New terms can be added.
- **Spatial Spread of Objects -** The spread of many static objects on a frame, that is characterized with terms as spread or concentrated. Adding and deleting description terms is possible through the corresponding buttons. The changes can be applied to all the channels in the respective frames. Also, for each term a confidence level and a text to save extra information about the term can be stored by double-clicking on the term, as shown in Figure 27.

Note that any change to the start and/or the end frame of the transition causes changes to the duration of the other shots/transitions, as described in Section *2.3.2.1. Shot Annotation.*

2.3.4.3. Key Segment Editing

By left clicking on a node which represents a key segment, the right part of the window displays the key segment's description.

According to Figure 29, the user can see and edit the following attributes:

- **ID** - The unique id of the key segment. The value cannot be changed.
- **Start Frame** - The first frame of the key segment.
- **End Frame** - The last frame of the key segment.

Note that any change can be applied to all the channels by checking the corresponding box, only if descriptions of the key segment exist in other channels.

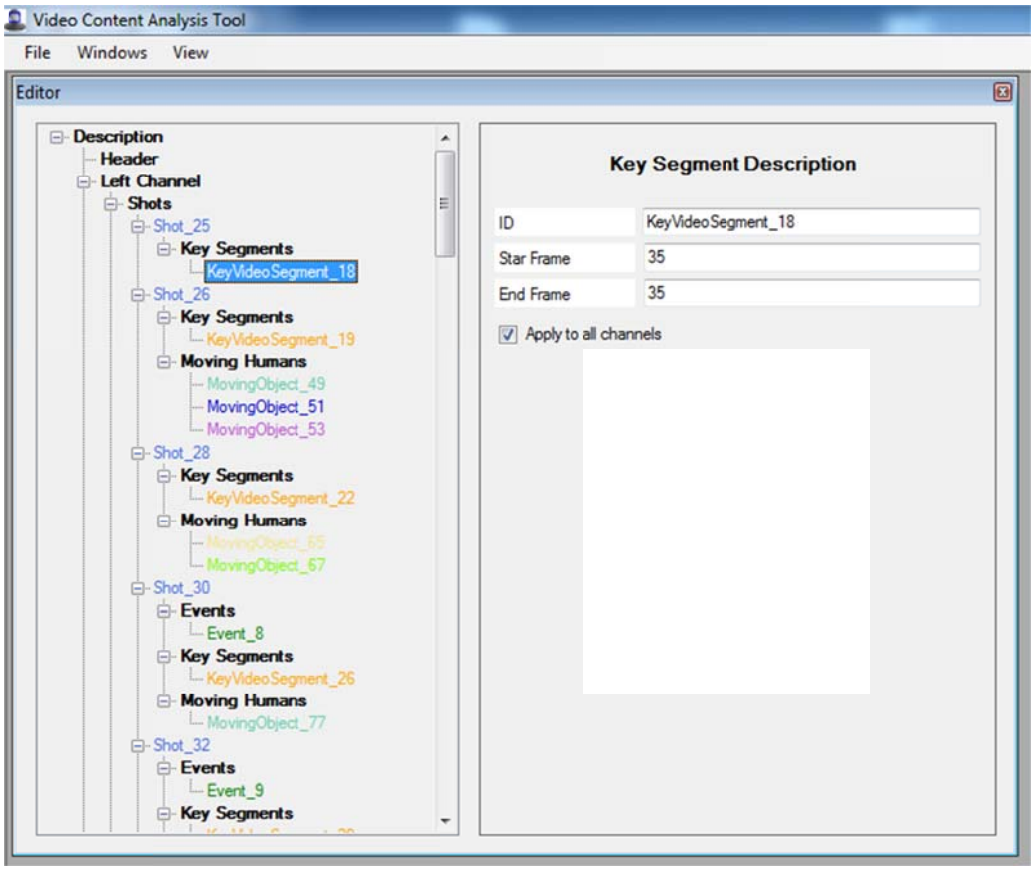


Figure 29: Key Segment Editing.

By right clicking on a node which represents a key segment, a dropdown menu will appear (Figure 30) through which the user can:

- Delete the description of the key segment.
- Delete the descriptions of the key segment from all the channels.
- Go to the description of the key segment in another channel.
- Copy the description of the key segment to another channel.
- Set as a description for a specific channel, an existing description of a key segment.

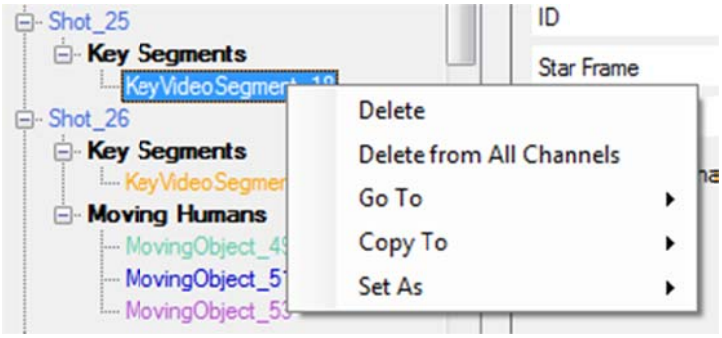


Figure 30: Right-clicking on a key segment node.

2.3.4.4. Event Editing

By left clicking on a node which represents an event, the right part of the window displays the event's description. So, according to Figure 31, the user can see and edit the following attributes:

- **ID** - The unique id of the event. The value cannot be changed.
- **Start Frame** - The first frame of the event.
- **End Frame** - The last frame of the event.
- **Event Type** - The type of the event. The value can be changed by selecting a new term from the corresponding drop-down list. New terms can be added.
- **Text** – A description of the event by using free text.

Note that any change can be applied to all the channels by checking the corresponding box, only if descriptions of the event exist in other channels.

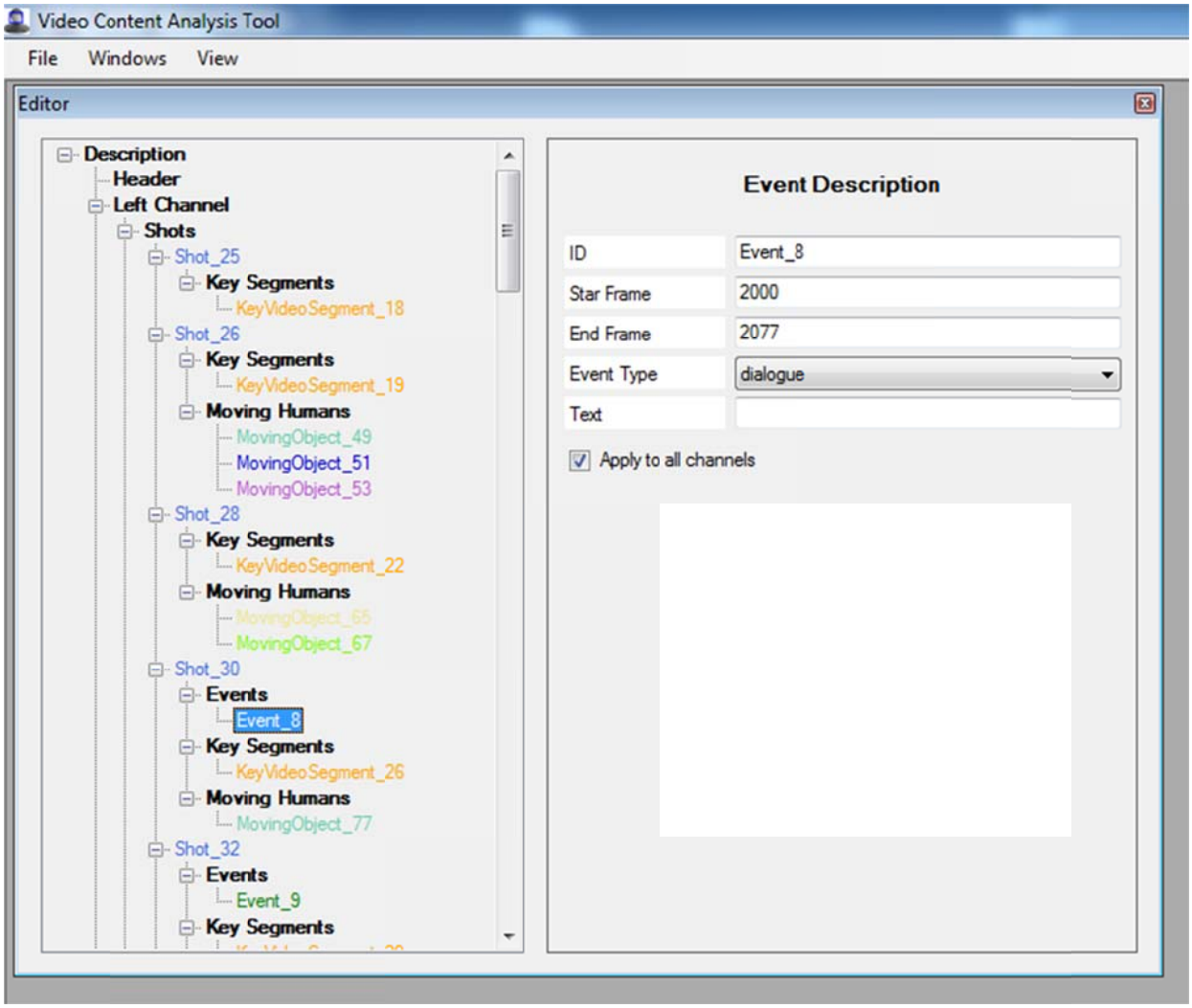


Figure 31: Event Editing.

By right clicking on a node which represents an event, a dropdown menu will appear (Figure 32) through which the user can:

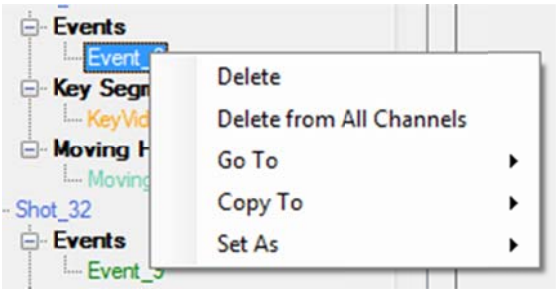


Figure 32: Right-clicking on an event node.

- Delete the description of the event.
- Delete the descriptions of the event from all the channels.
- Go to the description of the event in another channel.
- Copy the description of the event to another channel.
- Set as a description for a specific channel, an existing description of an event.

2.3.4.5. Static Object Editing

By left clicking on a node which represents a static object (i.e. an object whose appearance is marked on a single frame), the right part of the window displays the static object's description.

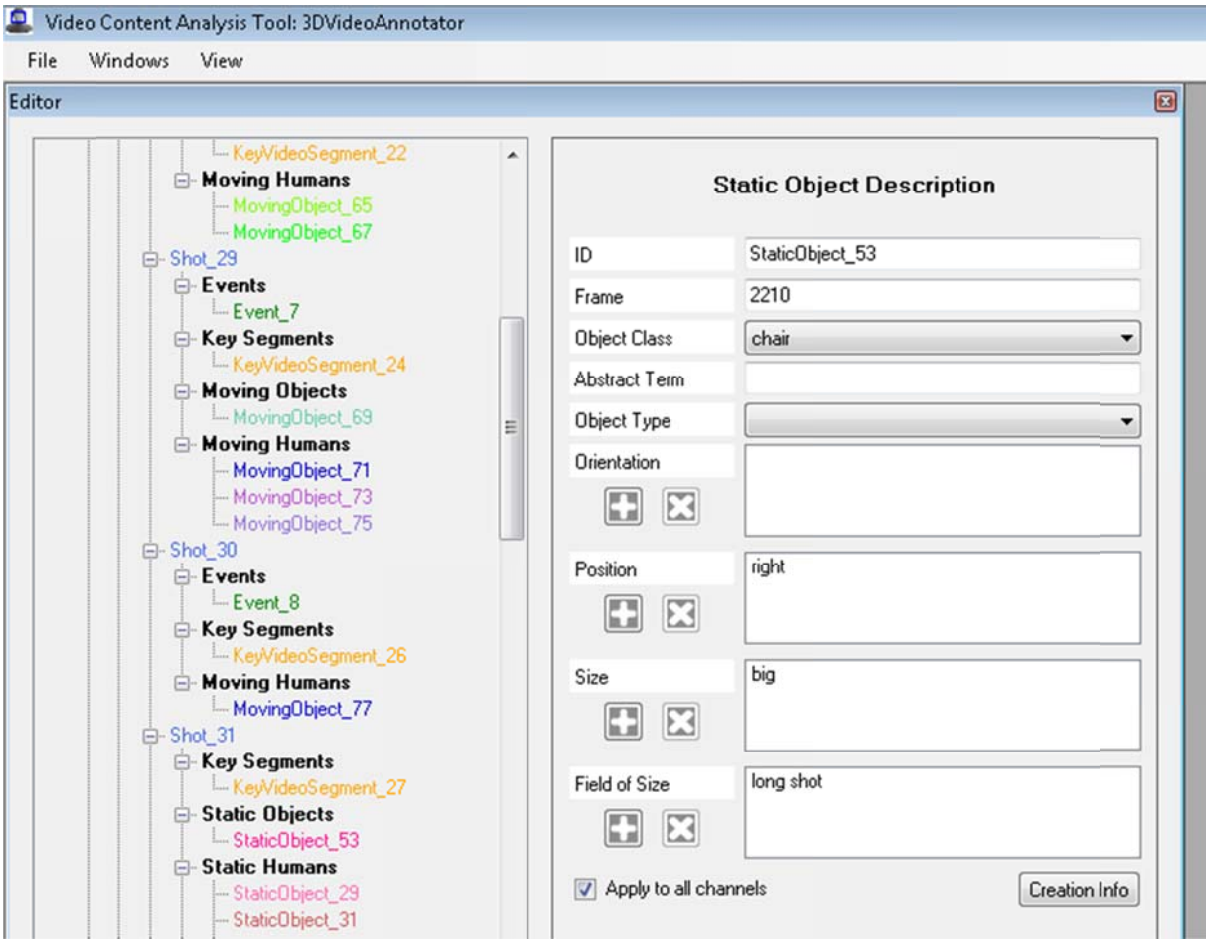


Figure 33: Static Object Editing.

According to Figure 33, the user can see and edit the following attributes:

- **ID** - The unique id of the static object. The value cannot be changed.
- **Frame** - The frame in which the static object appears. The value cannot be changed.
- **Object Class** - The object class, such as chair or car, is defined by selecting a term from the corresponding drop-down list. New terms can be added.
- **Object Type** - The specific type of the static object, for example an office chair. New terms can be added.
- **Orientation** - The orientation description of the static object, e.g., oriented left.
- **Position -** The position description of the static object, e.g., left.
- **Size -** The size description of the static object, e.g., small.
- **Size of Field** - The size-of-field description of the static object, e.g., close-up.

For the four last attributes adding and deleting description terms is possible through the corresponding buttons. Also, for each term a confidence level and a text to save some extra information about the term can be stored by double-clicking on the term, as shown in Figure 27.

Note that any change can be applied to all the channels by checking the corresponding box, only if descriptions of the static object exist in other channels.

By right clicking on a node which represents a static object, a dropdown menu will appear (Figure 34) through which the user can:

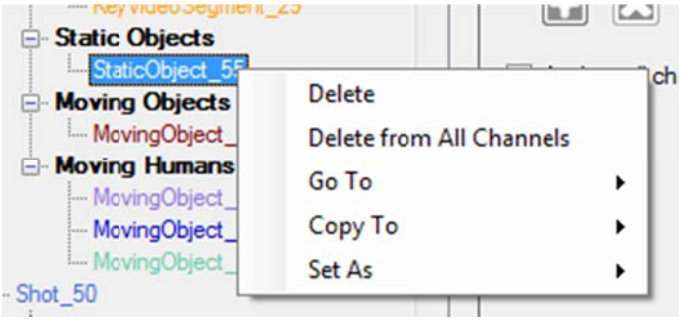


Figure 34: Right-clicking on a static object node.

- Delete the description of the static object.
- Delete the descriptions of the static object from all the channels.
- Go to the description of the static object in another channel.
- Copy the description of the static object to another channel.
- Set as a description for a specific channel, an existing description of a static object.

2.3.4.6. Static Human Editing

By left clicking on a node which represents a static human (i.e. a human whose appearance is marked on a single frame), the right part of the window displays the static human's description. So, according to Figure 35, the user can see and edit the following attributes:

- **ID** - The unique id of the static human. The value cannot be changed.
- **Frame** - The frame in which the static human appears. The value cannot be changed.
- **Body Part** - The body part of the human actor enclosed in the bounding box is defined by selecting a term from the corresponding drop-down list. New terms can be added.
- **Name** - The name of the human. This can refer to either an actual name (e.g., Bogart) or a symbolic name (e.g., person_1).
- **Activity** - The activity (e.g. walk) of the static human. The value can be changed by selecting a new term from the corresponding drop-down list. New terms can be added.
- **Expression** – The facial expression (e.g. anger) of the static human. The value can be changed by selecting a new term from the corresponding drop-down list. New terms can be added.
- **Orientation** - The orientation description of the static human, e.g., oriented left.
- **Position -** The position description of the static human, e.g., left.
- **Size -** The size description of the static human, e.g., small.
- **Size of Field** - The size-of-field description of the static human, e.g., close-up.

For the four last attributes adding and deleting description terms is possible through the corresponding buttons. Also, for each term a confidence level and a text to save some extra information about the term can be stored by double-clicking on the term, as shown in Figure 27.
Note that any change can be applied to all the channels by checking the corresponding box, only if descriptions of the static human exist in other channels.

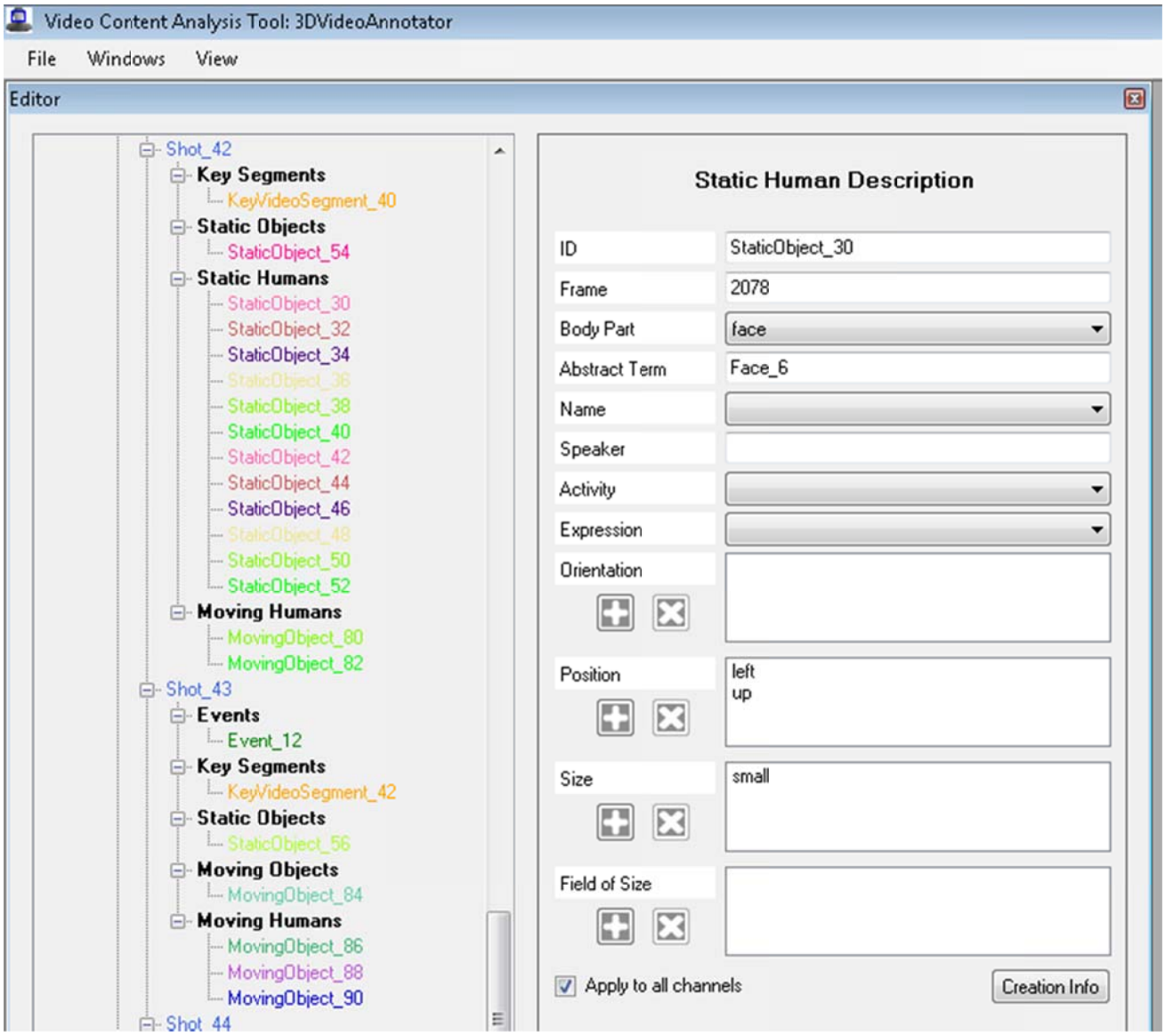


Figure 35: Static Human Editing.

By right clicking on a node which represents a static human, a dropdown menu will appear (Figure 36) through which the user can:

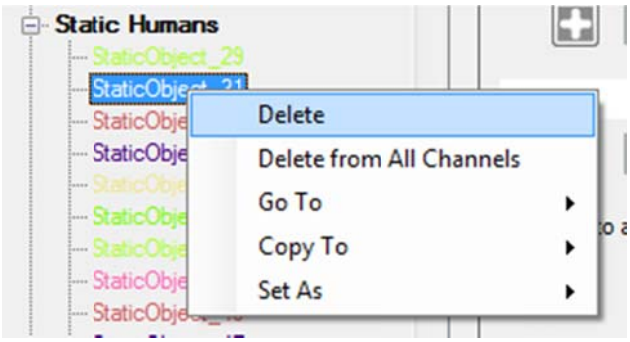


Figure 36: Right-clicking on a static human node.

- Delete the description of the static human.
- Delete the descriptions of the static human from all the channels.

- Go to the description of the static human in another channel.
- Copy the description of the static human to another channel.
- Set as a description for a specific channel, an existing description of a static human.

2.3.4.7. Moving Object Editing

By left clicking on a node which represents a moving object (namely a series of bounding boxes, over a number of consecutive frames that depict an object that moves over time), the right part of the window displays the moving object's description. So, according to Figure 37, the user can see and edit the following attributes:

- **ID** - The unique id of the moving object. The value cannot be changed.
- **Start Frame** - The start frame in which the moving object appears. The value cannot be changed.
- **End Frame** - The end frame in which the moving object appears. The value cannot be changed.
- **Object Class** - The object class, such as chair or car, is defined by selecting a term from the corresponding drop-down list. New terms can be added.
- **Object Type** - The specific type of the moving object, for example an office chair. New terms can be added.
- **Movement -** The movement of the moving object.
- **Position** - The position description of the moving object, e.g., left.
- **Size** - The size description of the moving object, e.g., small.
- **Size of Field** - The size-of-field description of the moving object, e.g., close-up.
- **Sub-Movement -** In case different movements occur within the same object appearance, the user can see and edit their durations and specific movements.
- **Related Movement –** The movement between this moving object and another moving object or human.

For the six last attributes adding and deleting description terms is possible through the corresponding buttons. Also, for each term a confidence level and a text to save some extra information about the term can be stored by double-clicking on the term, as shown in Figure 27.

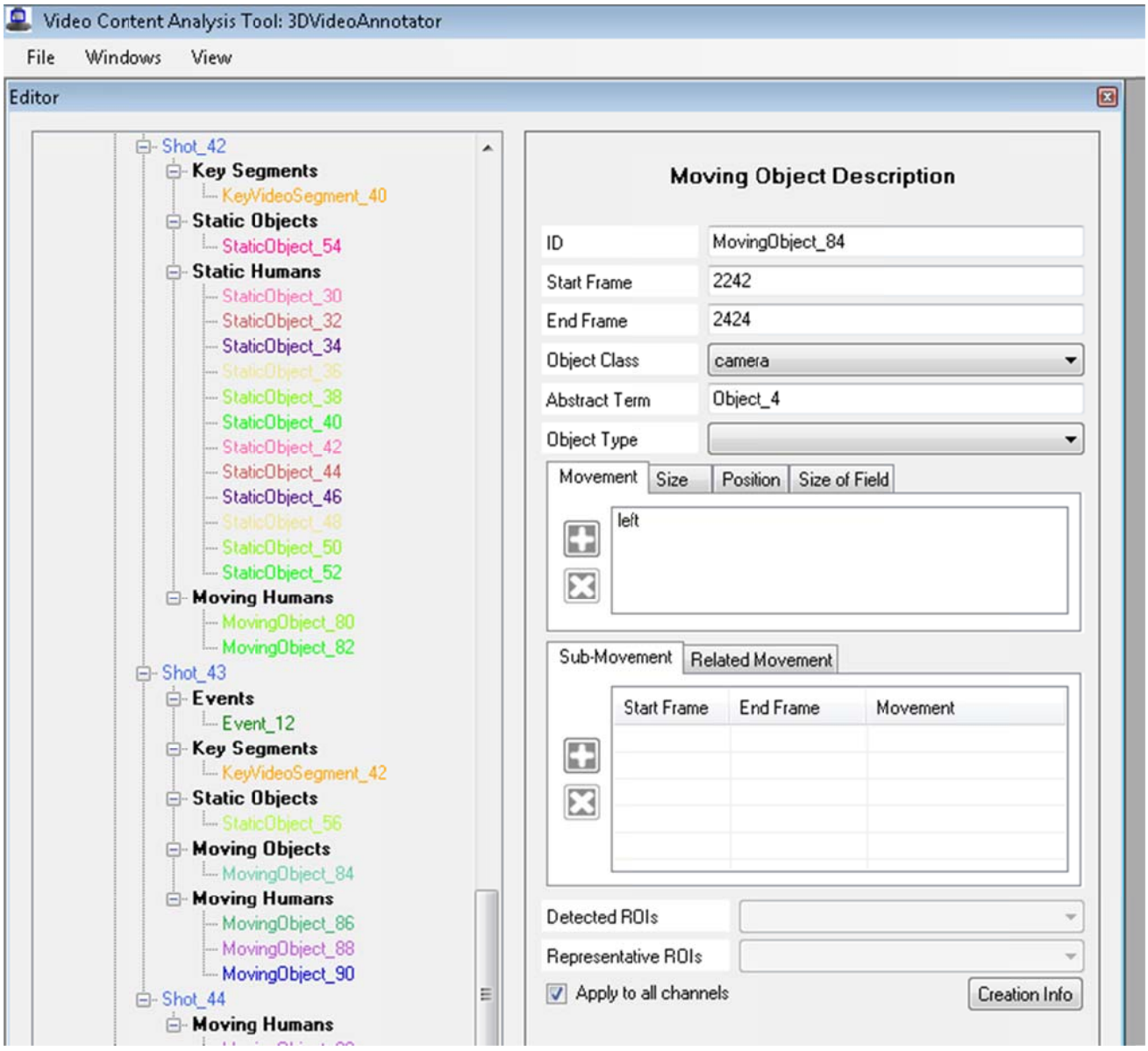


Figure 37: Moving Object Editing.

Note that any change can be applied to all the channels by checking the corresponding box, only if descriptions of the moving object exist in other channels.

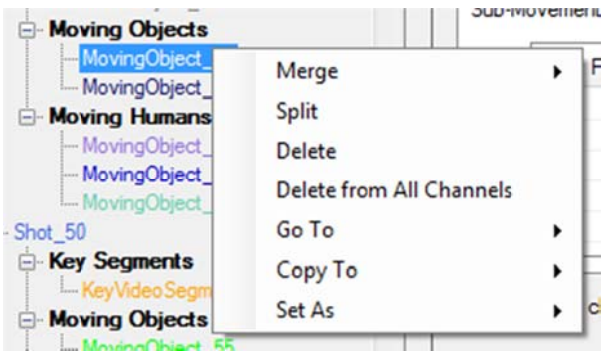


Figure 38: Right-clicking on a moving object node.

By right clicking on a node which represents a moving object, a dropdown menu will appear (Figure 38) through which the user can:

- Merge two moving objects, i.e. two set of bounding boxes (object trajectories). The two moving objects must have the same Object Class, have the same Object Type if such information is specified and appear in the same channels.
- Split the moving object into two moving objects.
- Delete the description of the moving object.
- Delete the descriptions of the moving object from all the channels.
- Go to the description of the moving object in another channel.
- Copy the description of the moving object to another channel.
- Set as a description for a specific channel, an existing description of a moving object.

2.3.4.8. Moving Human Editing

By left clicking on a node which represents a moving human (namely a series of bounding boxes, over a number of consecutive frames that depict a human that moves over time), the right part of the window displays the moving human's description. So, according to Figure 39, the user can see and edit the following attributes:

- **ID** - The unique id of the moving human. The value cannot be changed.
- **Start Frame** - The start frame in which the moving human appears. The value cannot be changed.
- **End Frame** - The end frame in which the moving human appears. The value cannot be changed.
- **Body Part** - The body part of the human actor enclosed in the bounding box is defined by selecting a term from the corresponding drop-down list. New terms can be added.
- **Name** - The name of the human. This can refer to either a actual name (e.g., Bogart) or a symbolic name (e.g., person_1).
- **Activity** - The activity (e.g. walk) of the static human. The value can be changed by selecting a new term from the corresponding drop-down list. New terms can be added.
- **Expression** - The facial expression (e.g. anger) of the static human. The value can be changed by selecting a new term from the corresponding drop-down list. New terms can be added.
- **Movement -** The movement of the moving human.
- **Position** - The position description of the moving human, e.g., left.

- **Size** - The size description of the moving human, e.g., small.
- **Size of Field** - The size-of-field description of the moving human, e.g., close-up.
- **Sub-Activity -** In case different activities occur within the same human appearance, the user can see and edit their durations and specific activities.
- **Sub-Expression -** In case different expressions occur within the same human appearance, the user can see and edit their durations and specific expressions.
- **Sub-Movement -** In case different movements occur within the same human appearance, the user can see and edit their durations and specific movements.
- **Related Movement –** The movement between this moving human and another moving object or human.

For the eight last attributes adding and deleting description terms is possible through the corresponding buttons. Also, for each term a confidence level and a text to save some extra information about the term can be stored by double-clicking on the term, as shown in Figure 27.

Note that any change can be applied to all the channels by checking the corresponding box, only if descriptions of the moving human exist in other channels.

By right clicking on a node which represents a moving human, a dropdown menu will appear (Figure 40) through which the user can:

- Merge two moving humans, i.e. two set of bounding boxes (human trajectories). The two moving humans must have the same Body Part, have the same Name if such information is specified and appear in the same channels.
- Split the moving human into two moving humans.
- Delete the description of the moving human.
- Delete the descriptions of the moving human from all the channels.
- Go to the description of the moving human in another channel.
- Copy the description of the moving human to another channel.
- Set as a description for a specific channel, an existing description of a moving human.

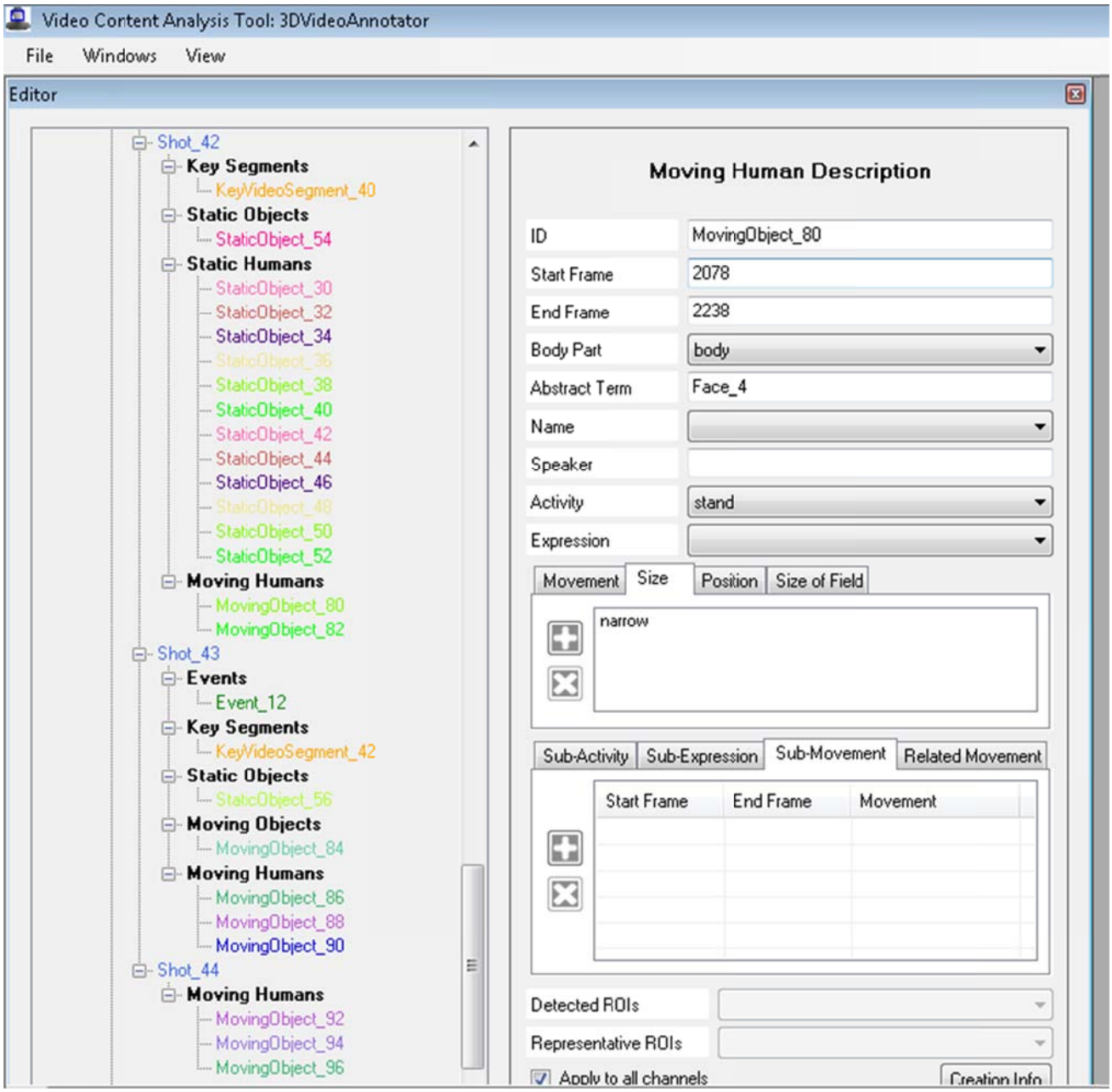


Figure 39: Moving Human Editing.

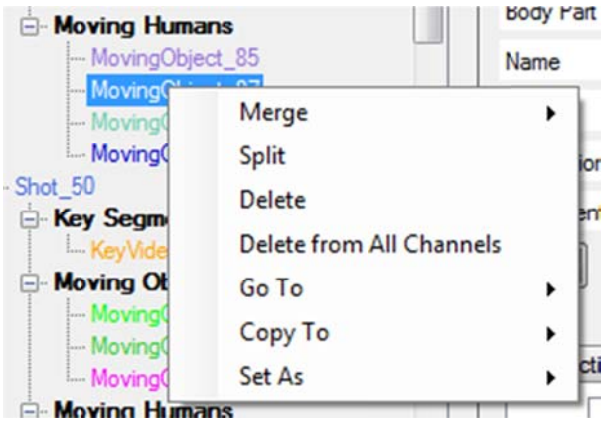


Figure 40: Right-clicking on a moving human node.

2.3.4.9. Cut Editing

By left clicking on a node which represents a cut, the right part of the window displays the cut's description. So, according to Figure 41, the user can see and edit the following attributes:

- **ID** - The unique id of the cut. The value cannot be changed.
- **Start Frame** - The first frame of the cut. The value cannot be changed.
- **End Frame** - The last frame of the cut. The value cannot be changed.
- **Characterization** - The cut can be characterized with terms, such as comfortable or uncomfortable for viewing, by selecting a characterization from the corresponding drop-down list. New terms can be added.

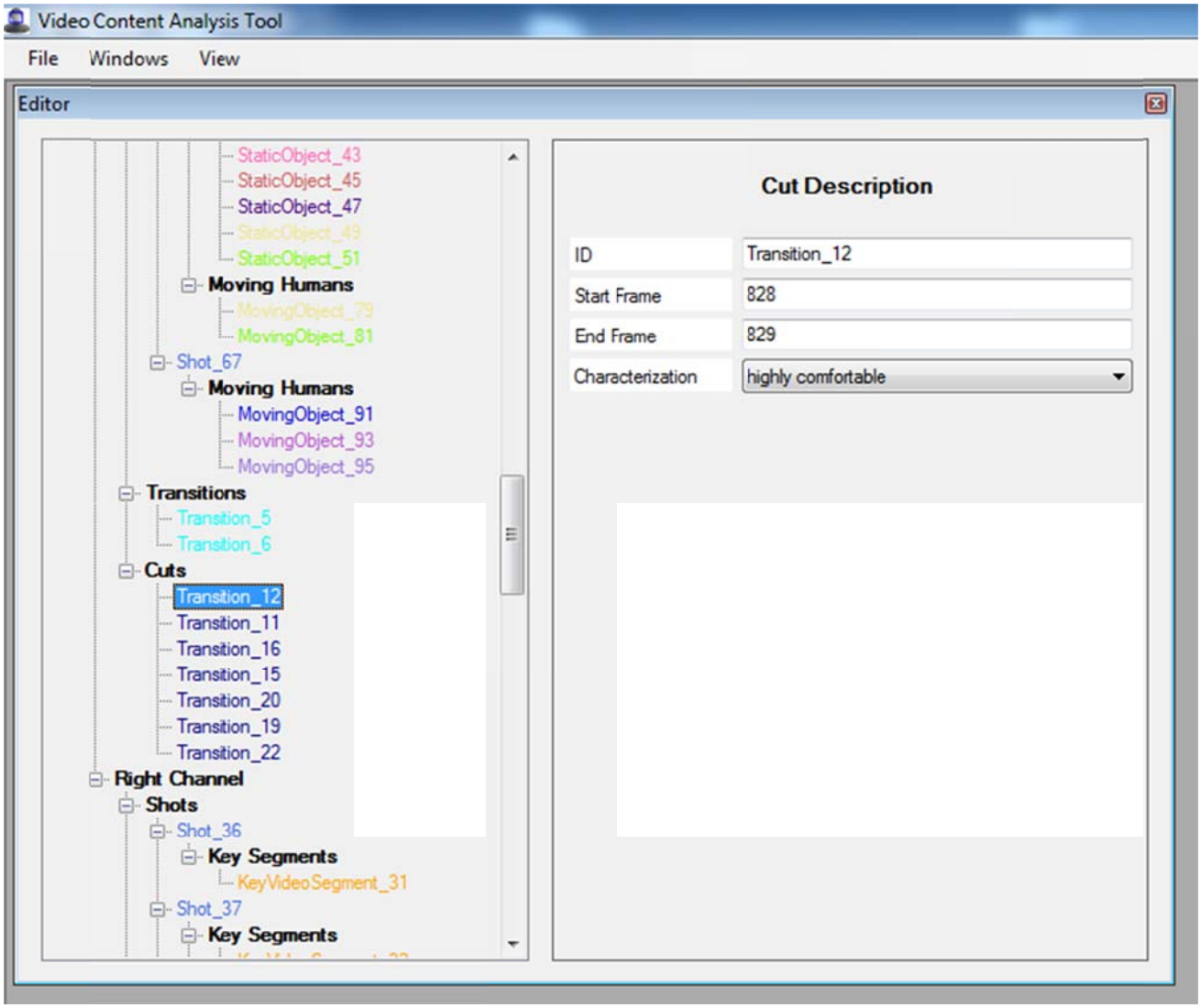


Figure 41: Cut Editing.

2.3.4.10. Header Editing

By left clicking on the node which is labeled “Header”, the right part of the window displays the “header” general information for the video, such as the location of the video, the compression, etc. So, according to Figure 41, the user can see and edit attributes regarding:

- The location and time of video production.
- The rights of the video content.
- The role and name of persons affiliated with the production (“Person” tabpage).
- Various parameters regarding
    - production (“Production Parameters” page)
    - video technical specifications (“Video Technical Information” page)
    - audio technical specifications (“Audio Technical Information” page)
    - specification of the subtitles (“Subtitles” page)
    - viewing conditions (“Monitor” page).

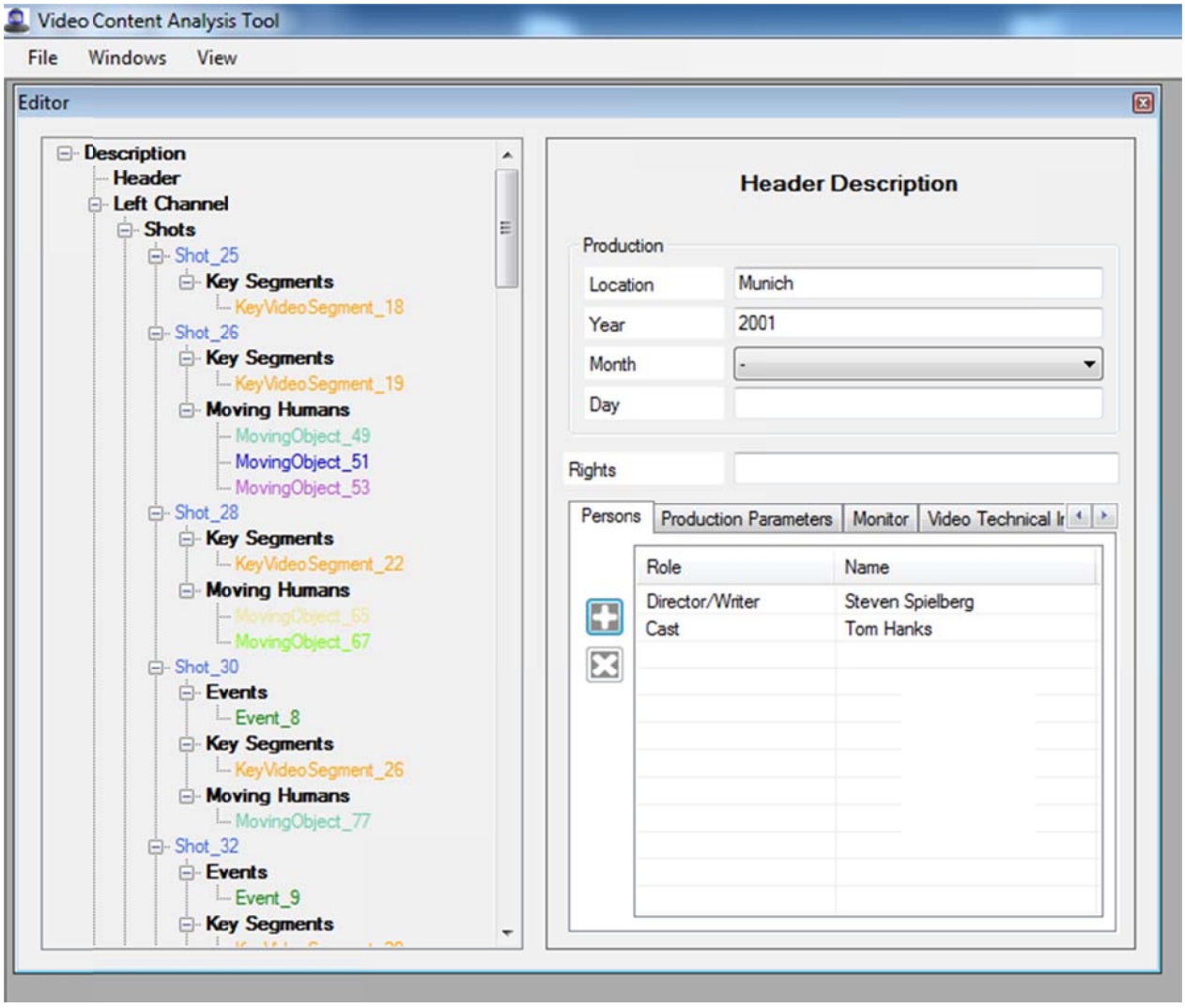


Figure 42: Header Editing.

2.3.5. Analyzer

The *Analyzer Window* (Figure 43) enables the user to execute various video analysis algorithms, such as face/body/object detection and tracking, shot detection, etc., by selecting a number of algorithms, defining its execution sequence and the video segments where the algorithms will be applied and setting which algorithms will be called in parallel and which ones are executed sequentially. When a group of algorithms is called in parallel, all the algorithms of the group are sequentially executed for each frame of the selected video segment. When algorithms are executed sequentially, an algorithm finishes processing of all frames of the segment and then the next one is executed.

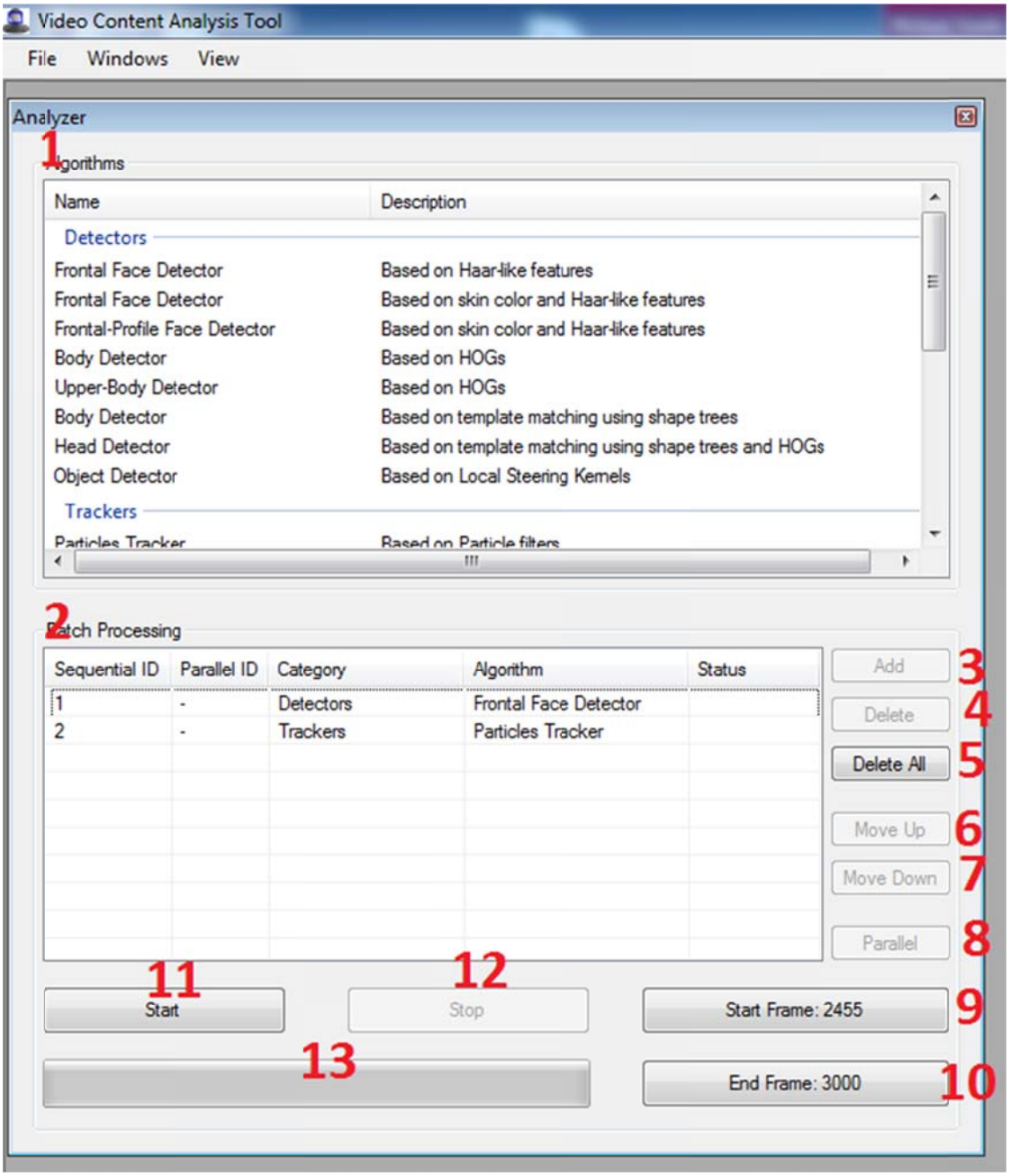


Figure 43: Analyzer Window.

A description of the various information areas and buttons of the Analyzer Window (Figure 43) is provided below:

1. Depicts the available algorithms. They are organized based on categories such as detectors, trackers, etc..
2. Depicts the selected algorithms, that will be executed on the selected video segment (*Batch Processing* list). The user can define their call sequence, delete them and set groups of algorithms which are called in parallel, through the corresponding buttons. See below.
3. It adds the currently selected algorithm from the *Algorithms* list to the *Batch Processing* list. An algorithm can be also added by double-clicking on it on the *Algorithms* list.
4. It deletes the currently selected algorithm from the *Batch Processing* list.
5. It deletes all the algorithms from the *Batch Processing* list.
6. It moves the currently selected algorithm up one slot.
7. It moves the currently selected algorithm down one slot.
8. It sets the currently selected algorithms to be executed in parallel.
9. It sets the first frame of the video segment where the algorithms will be applied.
10. It sets the last frame of the video segment where the algorithms will be applied.
11. It starts execution of the algorithms.
12. It stops execution of the algorithms.
13. It shows the progress of the batch processing.

The complete list containing the available algorithms is the following:

- A shot cut detector
- A key frame selector
- Three face detectors
- A tracker based on Particle filters
- A general object detector based on Local Steering Kernels (available only in the 32bit version)
- An object tracker based on Local Steering Kernels (available only in the 32bit version)
- An object tracker based on Local Steering Kernels (stereo version) (available only in the 32bit version)
- Three size-of-field characterization algorithms
- Two 3D quality defects detection algorithms (available only in the 32bit version)

a. In the following sections the corresponding manuals are given.

2.3.5.1. Shot Boundary Detector's Manual

2.3.5.1.1. Introduction

The Shot Boundary Detector is a software tool that provides users with the options to detect the shots in 3D videos. The algorithm uses Mutual Information for the detection of shots.

2.3.5.1.2. The parameter input menu

A parameter input menu (see Figure 4444) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

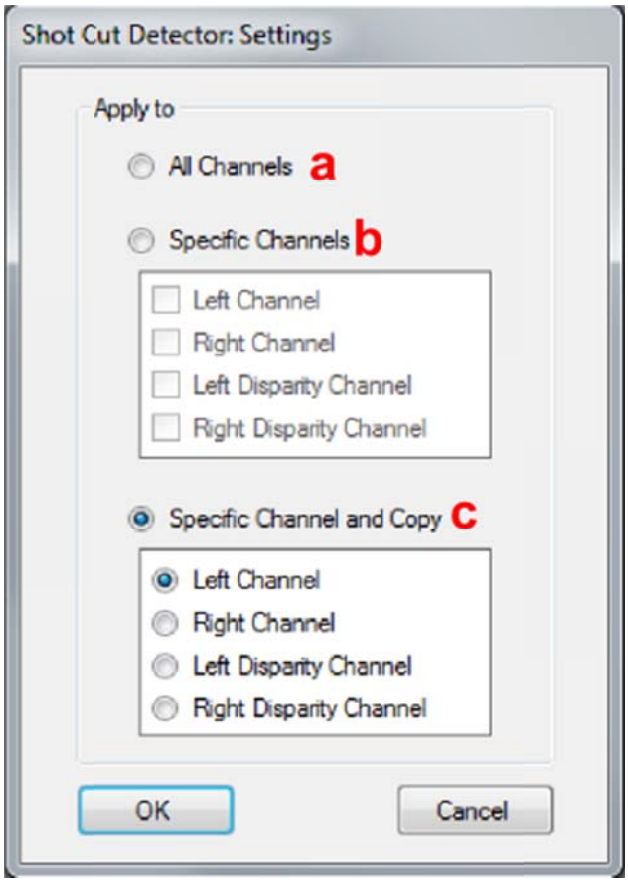


Figure 44: The parameter input menu

a. Option to apply shot detection to both channels.
b. Option to apply shot detection to either of the two video or disparity channels (left and/or right).
c. Option to apply shot detection to just one channel either of the two video or disparity maps (left or right) and transfer results to the other channel.

2.3.5.2. Haarcascade frontal face detector manual

2.3.5.2.1. Introduction

The Frontal Face Detector is an algorithm that provides users with the means to detect images of frontal faces from 2D/3D videos. The algorithm uses Haar-like features in order to calculate the frontal faces.

2.3.5.2.2. The parameter input menu

A parameter input menu (see Figure 35) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

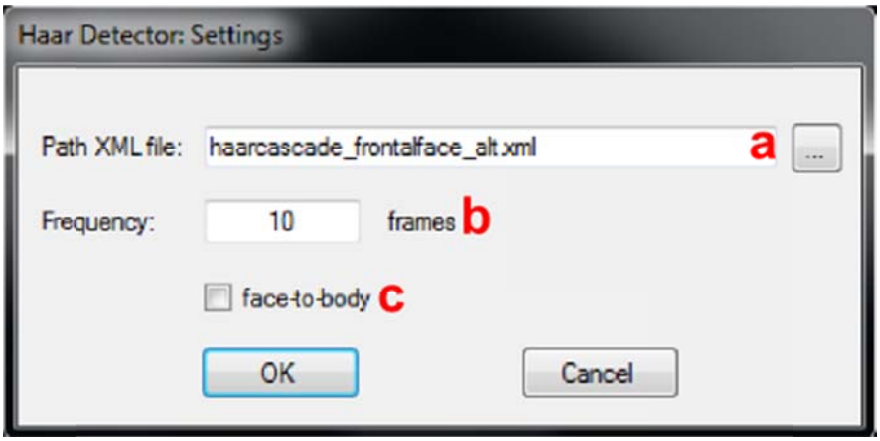


Figure 35: The parameter input menu

At the parameter input menu, the user can select:

a. The path to the xml file that contains the specifications for the Haar-like features.
b. The frequency of the face detection, **t**hat means how frequently (every how many frames) the Face Detector will be used (in the in-**b**etween frames faces are derived from the tracker used).
c. A face-to-body option that provides the possibility to return a ROI that contains also the body below the face that has been detected.

2.3.5.3. Color+Haarcascade Frontal Face Detector's Manual

2.3.5.3.1. Introduction

The Frontal Face Detector is an algorihm tool that provides users with the means to detect images of frontal faces from 2D/3D videos. The algorithm uses skin color in combination with Haar-like features in order to calculate the frontal faces. The skin color takes parameter in the HSV color space.

2.3.5.3.2. The parameter input menu

A parameter input menu (see Figure 46) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

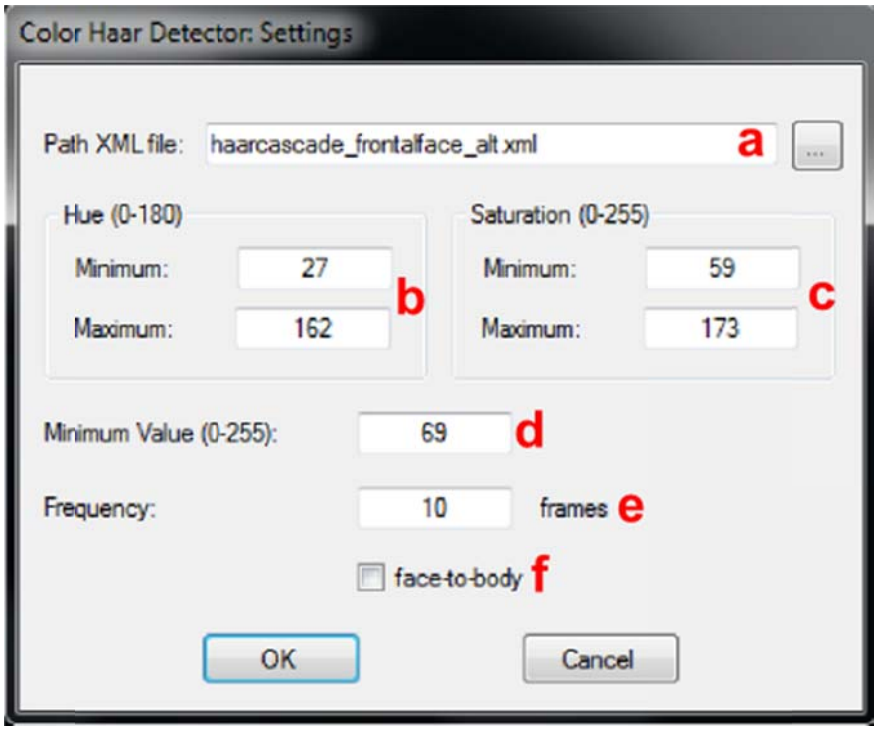


Figure 46: The parameter input menu

a. The path to the xml file that contains the specifications for the Haar-like features.
b. The minimum and maximum values for the Hue channel of the video being used. The range of the values (0-180) is shown on the GUI.
c. The minimum and maximum values for the Saturation channel of the video being used. The range of the values (0-255) is shown on the GUI.
d. The minimum value for the Value channel of the video being used. The range of the values (0-255) is shown on the GUI.
e. The frequency of the face detection, that means how frequently (every how many frames) the Face Detector will be used (in the in-between frames faces are derived from the tracker used).
f. A face-to-body option that provides the possibility to return a ROI that contains also the body below the face that has been detected.

2.3.5.4. Frontal–Profile Face Detector's Manual

2.3.5.4.1. Introduction

The Frontal–Profile Face Detector is a software tool that provides users with the means to detect images of frontal faces from 2D/3D videos. The algorithm uses skin color in combination with Haar-like features in order to calculate the frontal faces. The skin color takes parameter in the HSV color space.

2.3.5.4.2. The parameter input menu

A parameter input menu (see Figure) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

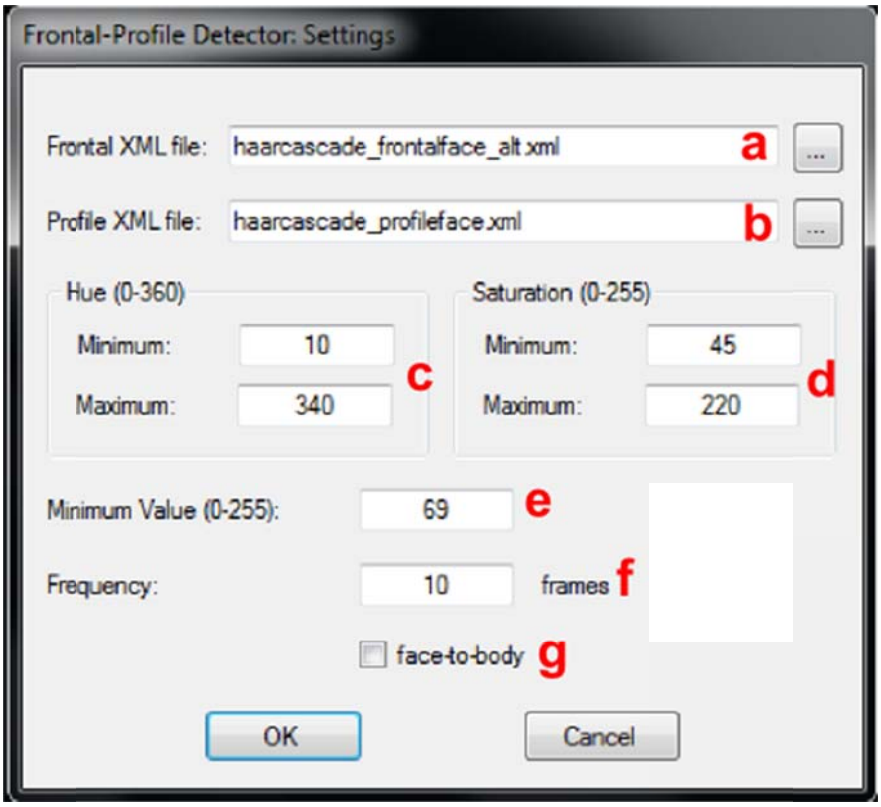


Figure 47: The parameter input menu

a. The path to the xml file that contains the specifications for the Haar-like features for the frontal facial image detection.
b. The path to the xml file that contains the specifications for the Haar-like features for the profile facial image detection.
c. The minimum and maximum values for the Hue channel of the video being used. The range of the values (0-180) is shown on the GUI.
d. The minimum and maximum values for the Saturation channel of the video being used. The range of the values (0-255) is shown on the GUI.
e. The minimum value for the Value channel of the video being used. The range of the values (0-255) is shown on the GUI.
f. The frequency of the face detection, that means how frequently (every how many frames) the Face Detector will be used (in the in-between frames faces are derived from the tracker used).
g. A face-to-body option that provides the possibility to return a ROI that contains also the body below the face that has been detected.

2.3.5.5. Object Detector's Manual

2.3.5.5.1. Introduction

The Object Detector is a software tool that provides users with the means to detect specified objects in 2D/3D videos. The algorithm uses Local Steering Kernels (LSKs) for the detection.

2.3.5.5.2. The parameter input menu

A parameter input menu (see Figure 48) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

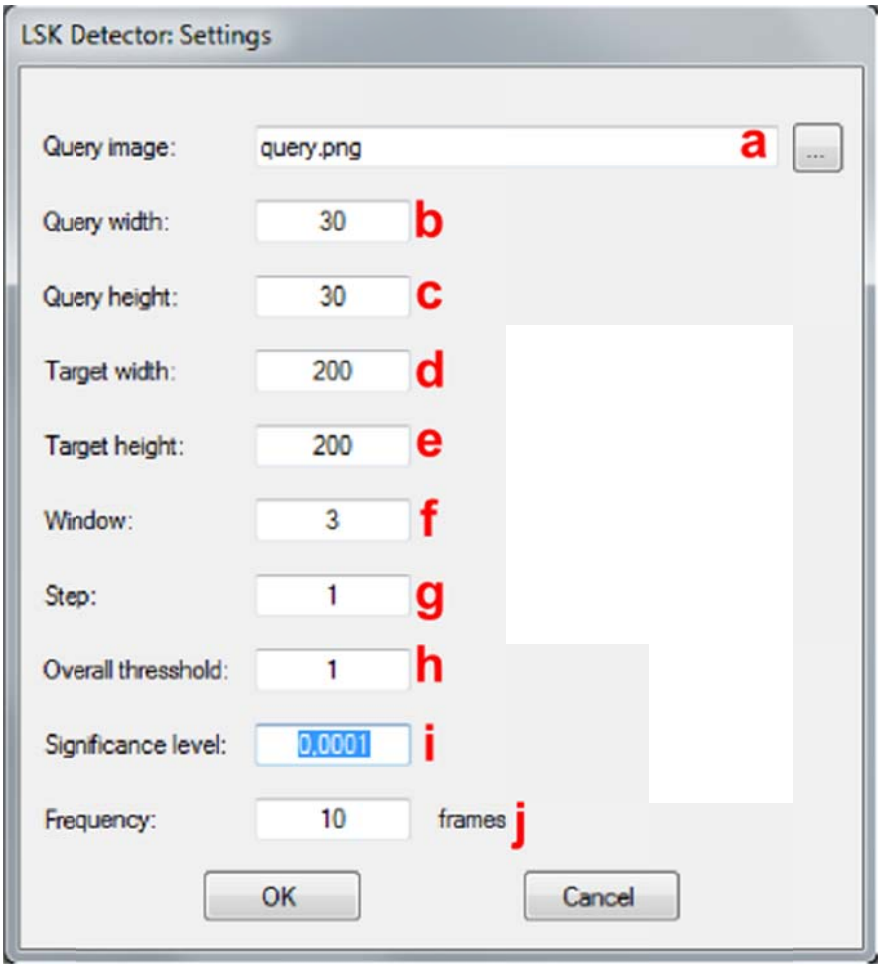


Figure 48: The parameter input menu

a. The path to the image file of the object to be searched for.
b. The image width to downscale the query image.
c. The image height to downscale the query image.
d. The image width to downscale the width of the video where the search is applied.
e. The image height to downscale the height of the video where the search is applied.
f. The window size of the LSK.
g. The step for the search (how thorough the search will be).
h. An overall threshold that is used to specify the existence of the object inside the frame.
i. A threshold that is used to specify the potential existence of more than one objects inside the frame.
j. The frequency of the object detection, that means how frequently (every how many frames) the Face Detector will be used (in the in-between frames faces are derived from the tracker used).

2.3.5.6. Particles Tracker's Manual

2.3.5.6.1. Introduction

The Particles Tracker is a software tool that provides users with the means to track an object on 2D/3D videos. The algorithm uses particles filters to track the object.

2.3.5.6.2. The parameter input menu

A parameter input menu (see Figure) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

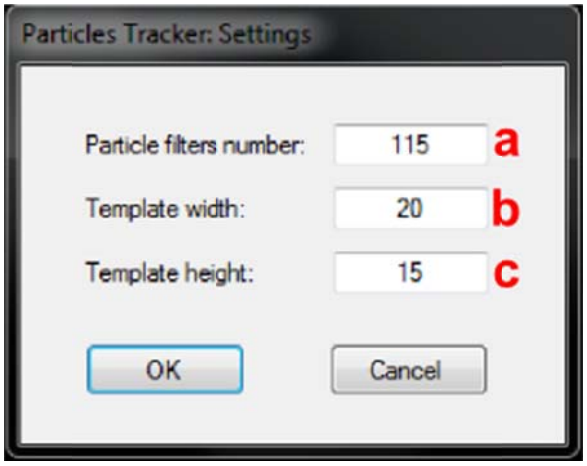


Figure 49: The parameter input menu

a. The number of particle filters that are going to be used.
b. The width of the downscaled image (template width).
c. The height of the downscaled image (template height).

2.3.5.7. LSK Stereo Tracker's Manual

2.3.5.7.1. Introduction

The LSK Stereo Tracker is a software tool that provides users with the means to track an object on 2D/3D videos. The algorithm uses Local Steering Kernels to track the object.

2.3.5.7.2. The parameter input menu

A parameter input menu (see Figure 50) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

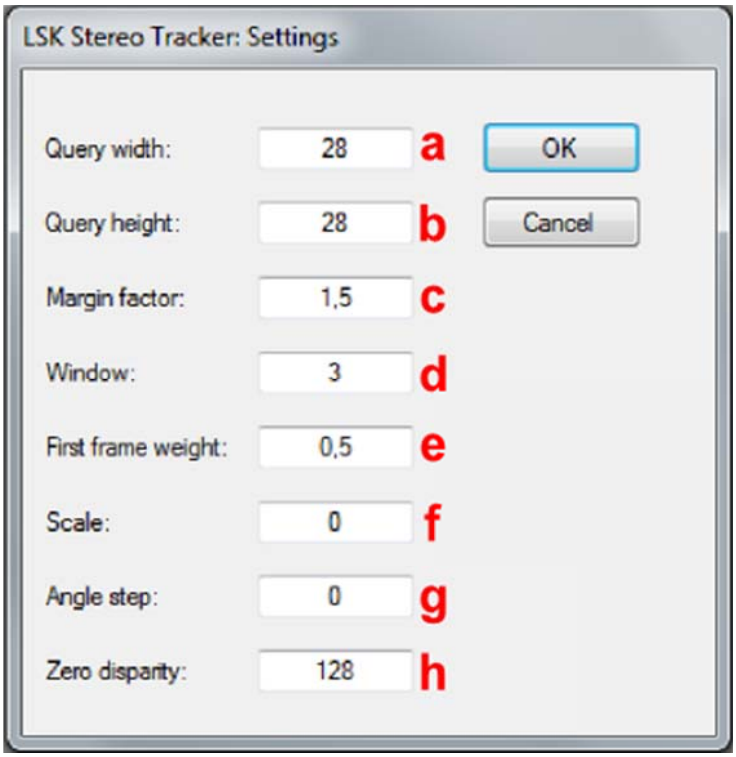


Figure 50: The parameter input menu

a. The image width to downscale the tracked object.
b. The image height to downscale the tracked object.
c. This option determines the size of search region.
d. This option determines the window size of the LSK.
e. Weight of the similarity with the object appearance in the first frame.
f. The scaling factor for the tracked image for the downscaled version of the tracked object.
g. The rotation factor (in degrees) for the tracked image for the rotated version of the tracked object.
h. The value for the zero disparity (on the screen neither in front nor behind the screen).

2.3.5.8. LSK Tracker's Manual

2.3.5.8.1. Introduction

The LSK Stereo Tracker is a software tool that provides users with the means to track an object on 2D/3D videos. The algorithm uses Local Steering Kernels to track the object.

2.3.5.8.2. The parameter input menu

A parameter input menu (see Figure 54) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

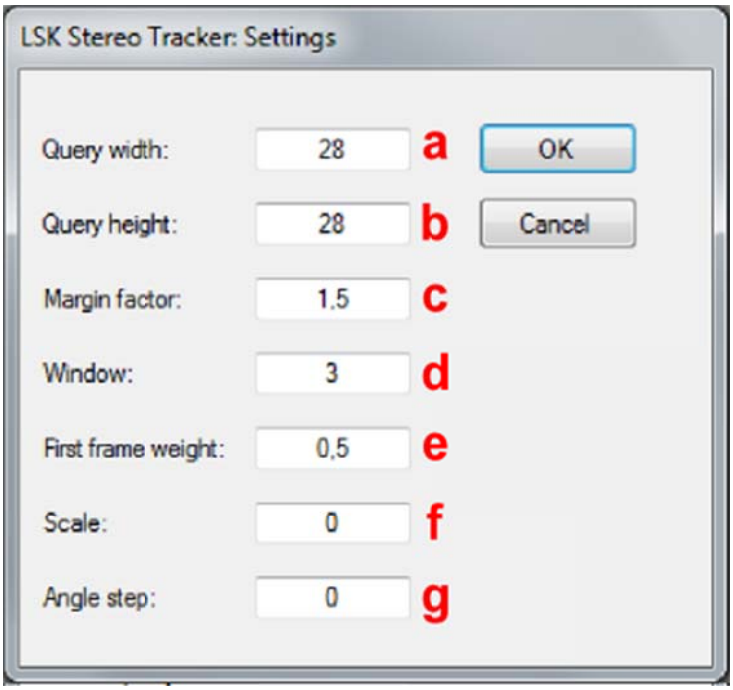


Figure 54: The parameter input menu

a. The image width to downscale the tracked object.
b. The image height to downscale the tracked object.
c. This option determines the size of search region.
d. This option determines the window size of the LSK.
e. Weight of the similarity with the object appearance in the first frame.
f. The scaling factor for the tracked image for the downscaled version of the tracked object.
g. The rotation factor (in degrees) for the tracked image for the rotated version of the tracked object.

2.3.5.9.3D Rules Detector's Manual

2.3.5.9.1. Introduction

The 3D Rules Detector is a software tool that provides users with the options to check 3D videos with disparity maps for violations of the 3D rules.

2.3.5.9.2. The parameter input menu

A parameter input menu (see Figure 52) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

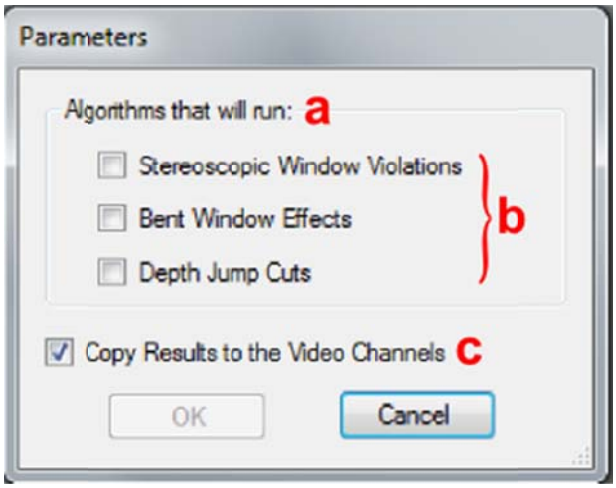


Figure 52: The parameter input menu

a. The algorithms on which test can be run are displayed in check buttons.
b. The options for algorithm include:
    I. Stereoscopic Window Violations,
    II. Bent Window Effects and
    III. Depth Jump Cuts
c. An extra option for marking the results in the video channels (normally the results are marked in the disparity maps only).

2.3.5.10. UFO Detector's Manual

2.3.5.10.1. Introduction

The UFO Detector is a software tool that provides users with the options to check 3D videos with disparity maps for object improperly displayed inside the theatre space (known as UFO).

2.3.5.10.2. The parameter input menu

A parameter input menu (see Figure) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

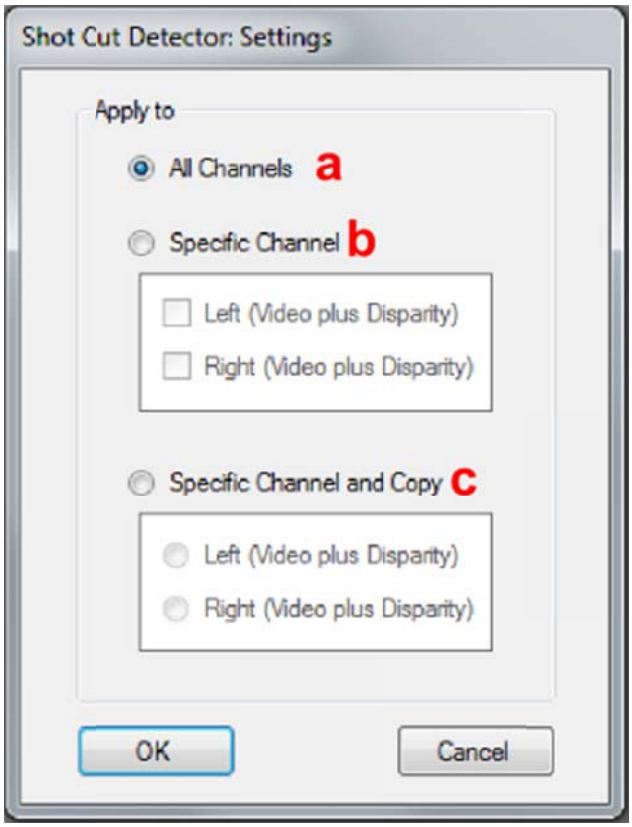


Figure 53: The parameter input menu

b. Option to apply the algorithm to both channels.
c. Option to apply the algorithm to either of the two channels (left or right) or even both.
d. Options to apply the algorithm to either of the two channels (left or right) and transfer the results to the other channel.

2.3.5.11. Keyframe Selection Tool's Manual

2.3.5.11.1. Introduction

The keyframe selector tool is a software tool that gives users the means to compute, visualize and manipulate keyframes of 2D/3D video shots. The tool is depicted in Figure 54.

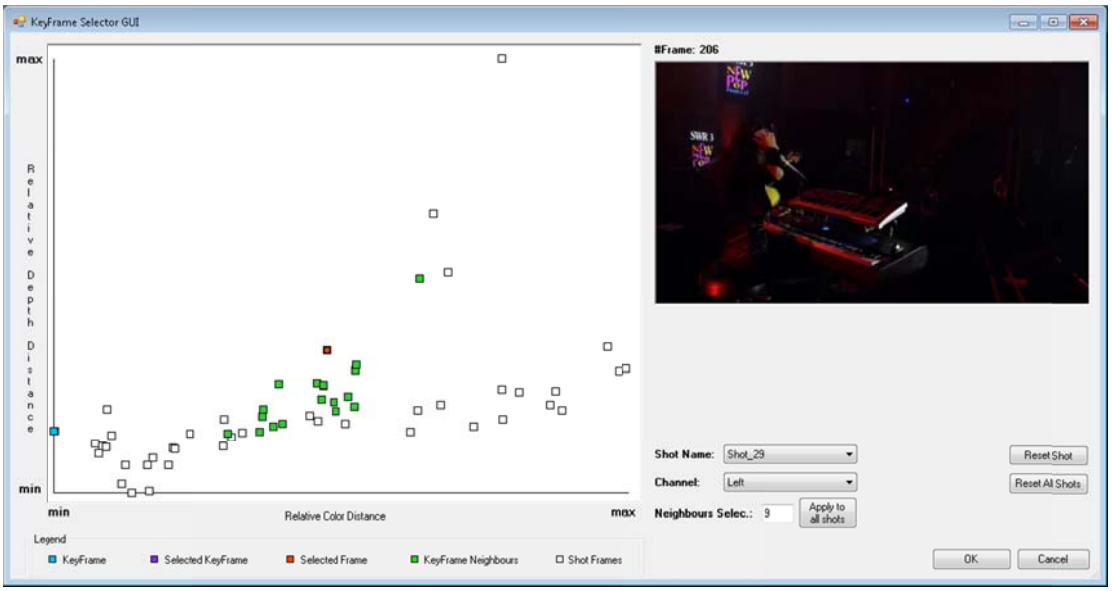


Figure 54: The keyframe selector GUI

At the time of this writing, three algorithm implementations are available, that can be selected from the parameter input menu. All three of them at their core need to compute distances between frames. For the first two algorithms, the distance between two frames is the sum of all their corresponding(having same coordinates) pixel distances. Pixel distances can be computed by two methods in this library:

a. **Distance of the averages of pixels:** initially an average value based on the RGB values of the pixel is computed (in essence the pixel is simply transformed to greyscale). The distance of two pixels is the distance of their average values.
b. **Euclidean distance of pixels:** the distance of two pixels is computed as an Euclidean distance, (the square root of the sum of the RGB values squared). This type of distance is a bit more precise but slower than the first one.

The algorithms in the input parameter menu are the following:

a. **Simple Distances of Frames:** This algorithm initially computes the distance for each shot frame pair (that is for frame pairs 1-2, 1-3, … , 2-3, …) where the distance between two frames is defined as the sum of their corresponding(having same coordinates) pixel distances, as mentioned above. After all distances among shot frames are computed, the keyframe can be derived as the one that has the smallest sum of frame distances, meaning that is the one closest to most other shot frames.
b. **Distances from Average Frame*:*** This algorithm computes an "average" shot frame, which in essence is a frame whose pixels hold the average value of all the shot frames' corresponding pixels. The keyframe will then be the one that has the least distance from the average frame. Frame distances are also computed based on pixel distances here, as in the first algorithm. This is by far the fastest algorithm of the two but slightly less accurate.
c. **Distances of frame Histograms*:*** This algorithm follows a similar process to KFSelectorAllDistances to produce its keyframes, with the only difference being that the distance between two frames in this algorithm is not the sum of their corresponding pixel

distances, but the distance of their histograms. This algorithm is the most context sensitive of the three, and can yield drastically different results from the first two. For histogram distances, the metrics provided by OpenCV are used as is: *Correlation*, *Chi-Square*, *Intersection*, *Bhattacharyya*.

2.3.5.11.2. The parameter input menu

A parameter input menu (see Figure 55) appears when the user starts the tool from the Video Content Analysis Tool's analyzer window.

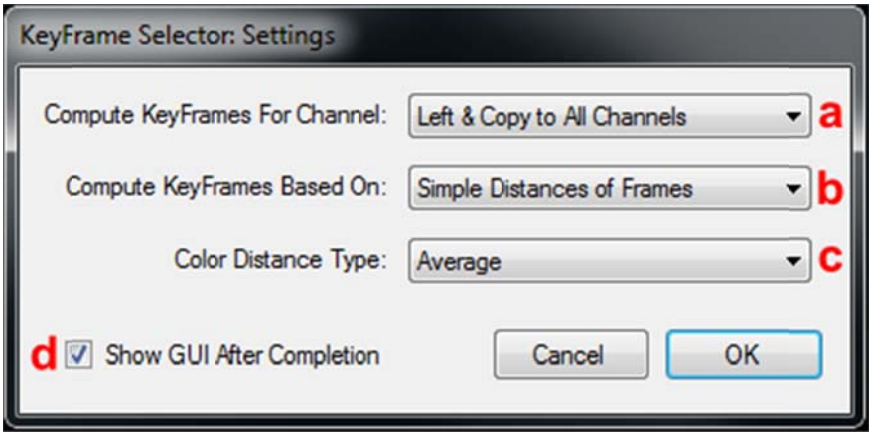


Figure 55: The parameter input menu

At the parameter input menu, the user can select:

d. The channel on which the keyframe selection will be based (for a 3D video) and if she wants the resulted keyframes to be stored to all channels in the Video Content Analysis Tool.
e. The algorithm that will be used for keyframe computation. The algorithms available.
f. The algorithm dependent distance type the algorithm will use to compute frame distances.
g. If the GUI should appear after the computation of the keyframes or not. If this is not checked, the results will be immediately stored in the "Video Content Analysis Tool", without any changes, and neighbouring frames will be stored along with the keyframe, forming a key-segment of the shot instead. The key-segment will consist of 21 frames, with the keyframe in the middle of the segment.

2.3.5.11.3. Result representation and manipulation

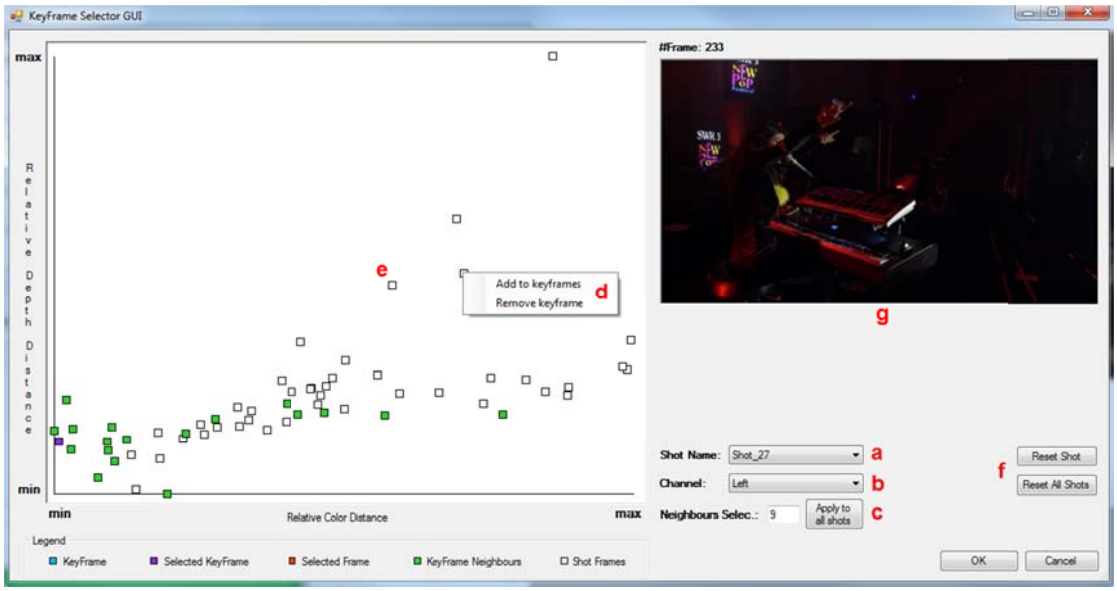


Figure 56: The parameter input menu

If the user has checked the option marked as “d” in Figure 55, the GUI will appear after the keyframe computations (see Figure 56).

The white panel on the left of the GUI contains the shot frames, which are represented by small squares; they can be clicked and visualized in the upper right corner of the GUI (see Figure 54). This panel basically is a graphical representation of the selected shot’s frames success as representative frames. The most valuable representative frames, are placed closer to the start of the axes. The horizontal axis represents the aggregate color distance of the shot frame; that is the sum of frame distances the specific frame has to all the rest of the shot. Similarly, the vertical axis represents the aggregate depth distance, if disparity videos are also available. On the far left side of the panel, the frame with the smallest aggregate color distance is placed, which means it has the biggest similarity of color to most other shot frames, thus being the best shot representative between the colored frames. Similarly, the rightmost placed frame will be the worst shot representative between the colored frames, an outlier. The same logic holds for distances that are placed on the vertical(depth) axis. Lastly, color and depth aggregate distances are combined (through an Euclidean distance metric) and yield the best candidate keyframe.
Each frame’s actual aggregate distance values can be explicitly seen when the mouse pointer hovers over its square; a context menu displaying them appears, as can be seen in Figure 57.

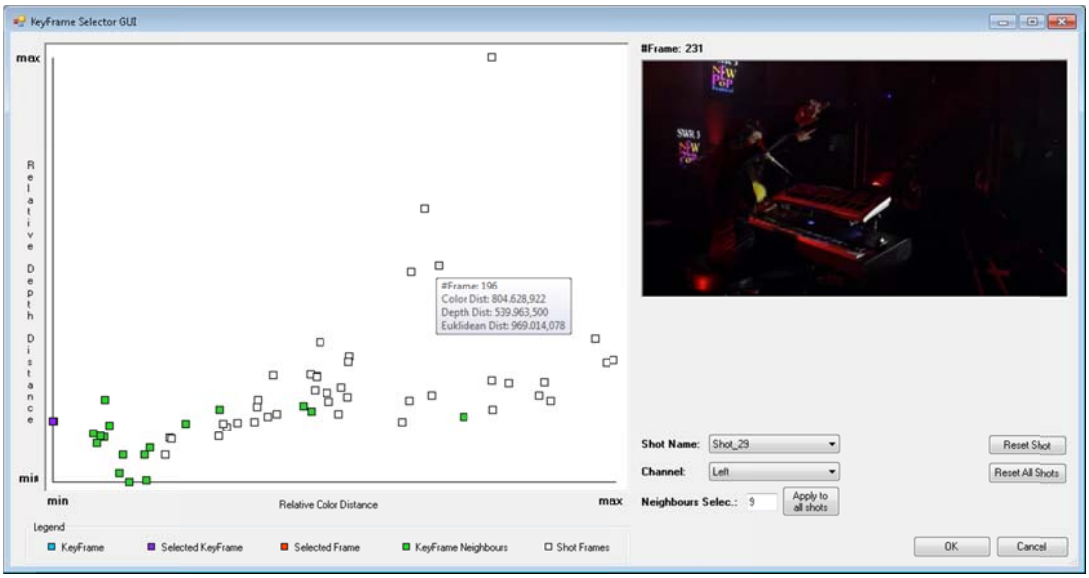
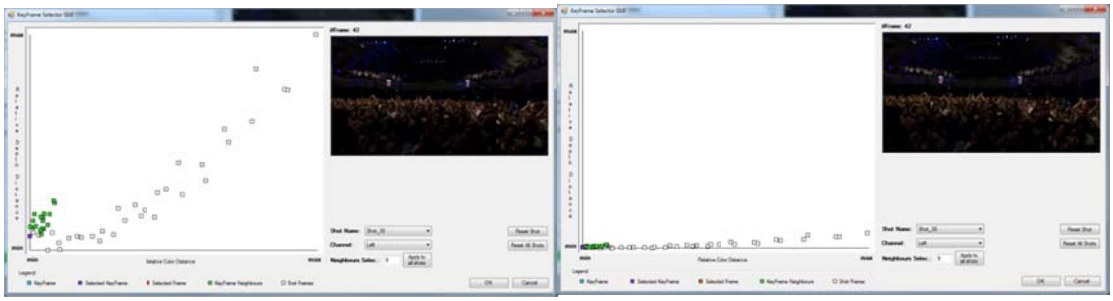


Figure 57: The context menu that appears when hovering over a frame's square.

For user convenience, the graph is stretched on both directions so that frames can be easily clicked on the graph and not become overly congested. This feature has been implemented, because the largest aggregate distance on one axis can by far outweigh the other largest aggregate distance, with the resulting frame squares becoming greatly congested (see Figure 58-b). The proportions can change, by right-clicking somewhere in the white space of the graph and selecting "*Switch between real/stretched proportions*". The difference of the two modes can be seen in Figure 58. The real proportions graph can be useful for the user to determine how much more contribution color had over depth (and vice versa) for the keyframe computation.



A **B**

Figure 58: a. Stretched Graph, b. Real Proportions Graph

The user can also easily add keyframes to the shot or remove them, by right-clicking on the respective square, as seen in Figure 59.

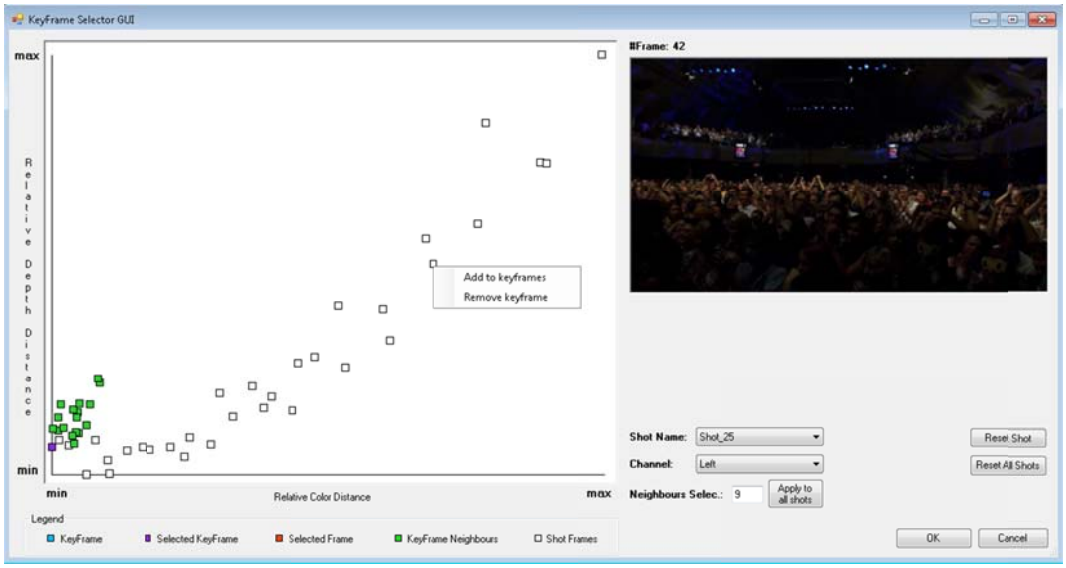


Figure 59: The context menu that appears on right clicking a frame square.

If a small trailer-like video of the original one is to be made out of keyframes, a single keyframe for each shot is not enough. For that reason, some neighbouring frames can also be attached to the keyframe, with all together forming a key-segment of the shot. The GUI gives the user the ability to select the number of neighbours that will be attached to the keyframe from its left and right independently, that is, the final key-segment will consist of numOfNeighbours + 1 + numOfNeighbours frames. A frame's neighbours can also be seen on the graph (their squares are green-colored, as mentioned later), giving the user the opportunity to view how a keyframe's neighbouring frames relate to it.

A color scheme has been implemented for the graph's frame squares to make important frames on the graph distinguishable (see Figure 54):

- keyframe squares are blue-colored, unless they are clicked by the user, which makes them purple
- a clicked frame square is red-colored, unless it is a keyframes as mentioned above
- the clicked frame's neighbouring frames are green-colored (their number can be changed on the GUI).

If the user needs to revert changes, he can click the appropriate reset buttons located on the bottom right part of the GUI (see Figure 60). All changes made be the user will be reverted.

Last but not least, the shot (or channel for a stereo video) can be changed from the bottom right corner of the GUI (see Figure 60). When this happens, the GUI removes old frames from the graph and updates it with the new selected shot's frames, with the frame depicted on the upper right of the GUI changing to a keyframe of the new shot automatically.

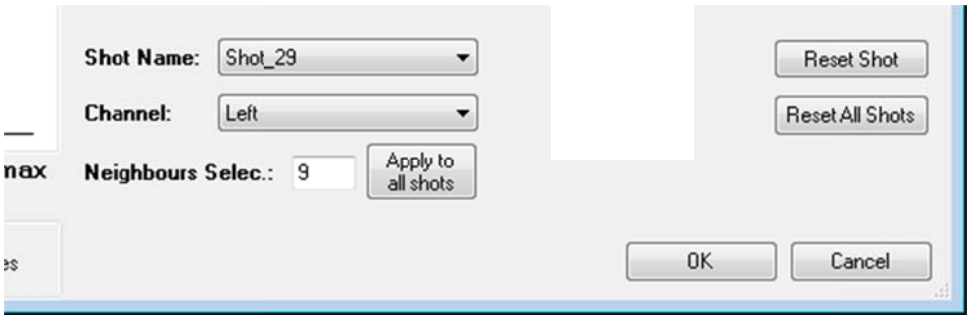


Figure 60: The bottom right corner of the GUI.