

Dialogue Detection

Version 1.1

The Dialogue Detection module is used to detect dialogue, monologue or no speech in a video stream based on Lip Activity of each frame and the number of Actors that participate in the speech. To install the Dialogue Detection module for DIVA3D, the user has to copy the DialogueDetection.dll in DIVA3D working directory where DIVA3D.EXE resides. The module will be loaded automatically by DIVA3D the next time it is started and the entry *Dialogue Detection* will be displayed (Figure 1).

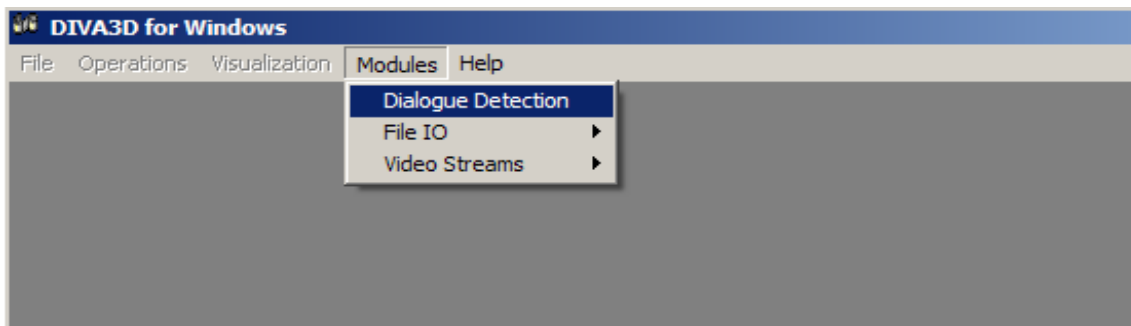


Figure 1: A view of the Dialogue Detection module submenu.

Usage

To use the Dialogue Detection module, first import an input video stream to Diva3D using *File IO -> Import* Module. In this point, it should be mentioned that this input video stream should contain only one actor per frame (or it could be a frame without actors – detections) due to the fact that the information produced by the Visual Speech Detection refers to a video stream with only one actor per frame or frames without face detections. After this, by selecting Dialogue Detection from the Modules Menu, the user is prompted with a dialog box to select a video stream to be processed.

The next step is typical the file open dialog box so that the user can select the XML file produced by the Visual Speech Detection module containing the information about the Lip Activity for each frame with detection of the selected video file (1 if there is lip activity and 0 if not).

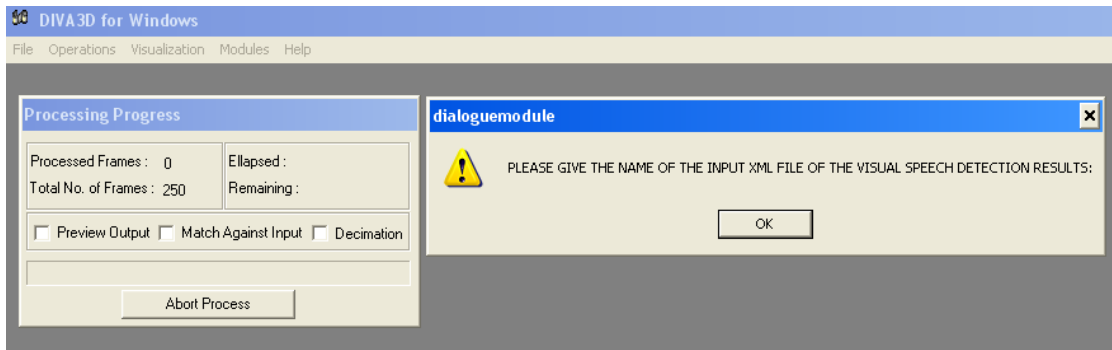


Figure 2: After pressing the “ok “ button the user can provide the input XML file from the Visual Speech Detection module.

After the opening of the XML file from the Visual Speech Detection module the user has to select the XML file produced by the Face Clustering module, which contains which Actor is present in each frame with detection.

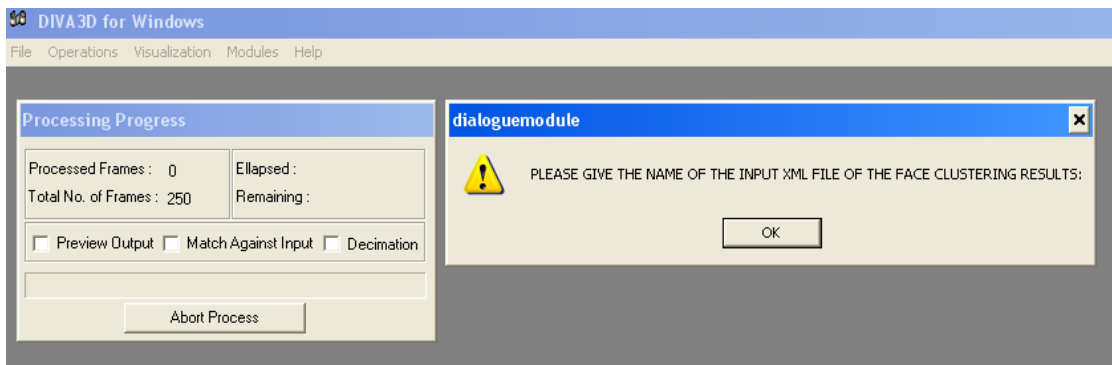


Figure 3: After pressing the “ok “ button the user can provide the input XML file from the Face Clustering module.

The lip activity produced by the Visual Speech Detection module stores the speaking information for each actor appearance. In the same way, the results of the Dialogue Detection module represent the dialogue or no dialogue decision for each actor appearance. The algorithm finds segments of lip activity for each actor and then examines every segment with the segments of other actors. If one segment overlaps with another segment of a different actor, then the module decides that there is dialogue. The same result occurs in case two segments of lip activity of different actors have a gap smaller than 25 frames. In other cases, such as only one actor have frames with lip activity or there is a gap of 40 frames between two speaking segments, the result is a monologue. The last possible result is “No_Dialogue” and it is produced when there are actors with no speaking segments.

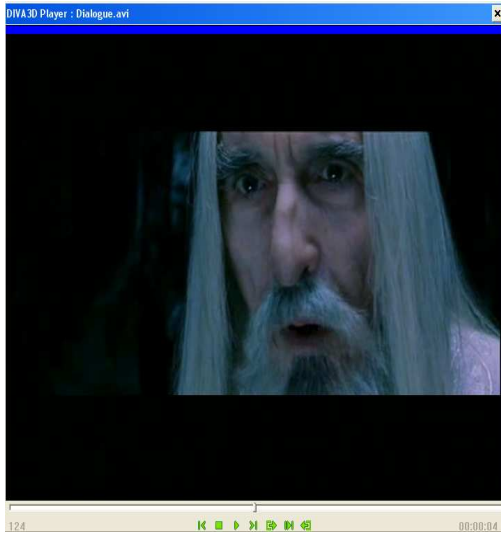


Figure 4: Dialogue scene (blue stripe).
(red stripe).



Figure 5: Non Dialogue scene

The results of the Dialogue Detection module are saved in XML format. The dialogue information is saved in this way : each actor appearance has value "Dialogue" if dialogue is detected, "Monologue", if the actor in this actor appearance speaks but the other actor appearances in its "temporal" neighborhood are silent, and last, "No_Dialogue" is detected if there is not leap activity in the actor appearance.