



Theme 1

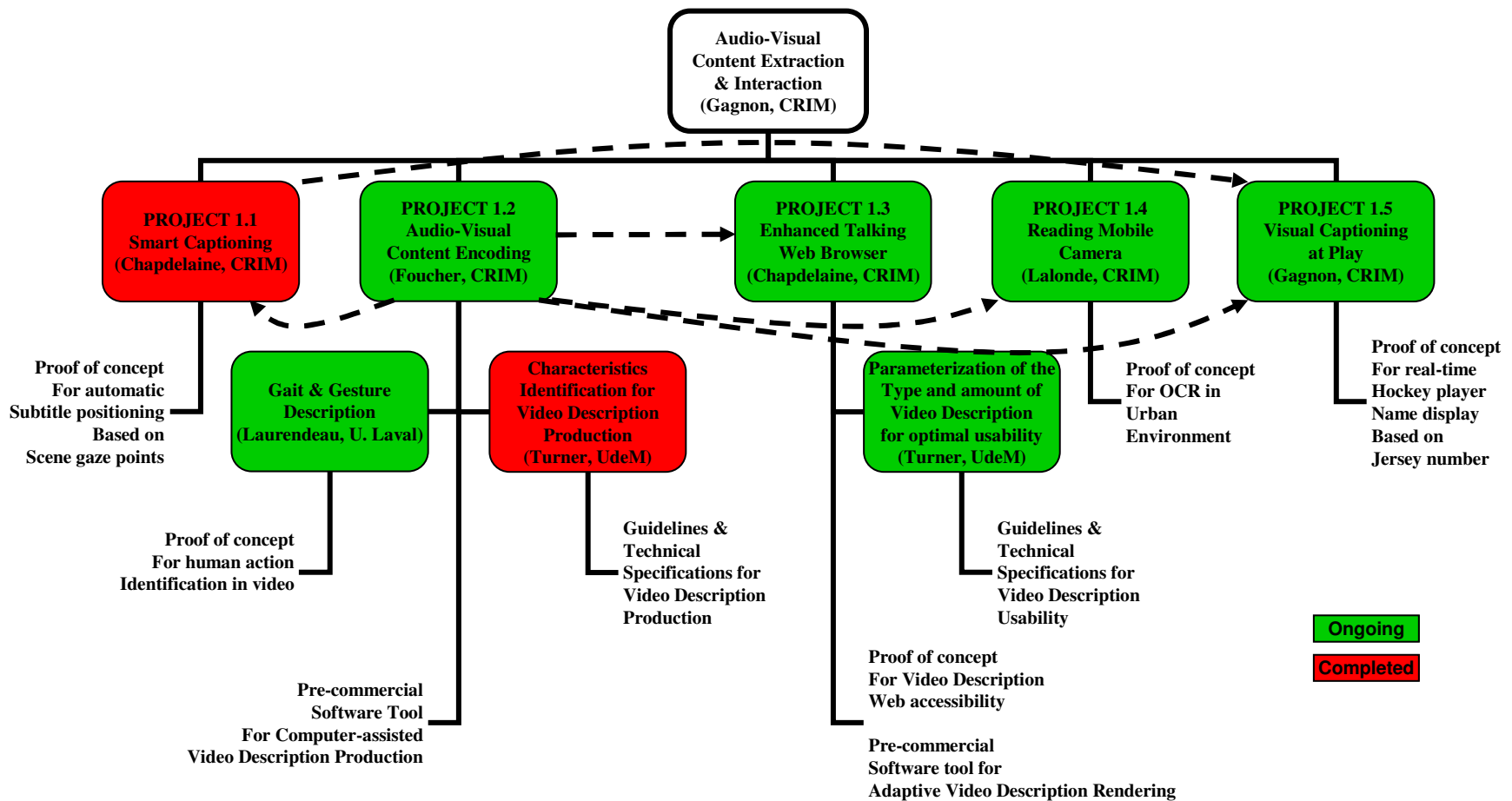


Audio-visual content extraction & interaction



Langis Gagnon, PhD
Theme leader
Vision & Imaging Team
CRIM

Theme 1 Projects Map





Projects 1.2 & 1.3 *Goals*

- **Computer-vision software tools** to help production and accessibility of **Video Description**

- Software tools
 - **Video Description manager (beta version)**
 - For managing audio-visual content extraction modules (shot, faces, text, motion, places, etc.) and generating VD drafts
 - **Video Ground Truth maker (beta version)**
 - For labeling video content for performance measures of automatic indexing tools
 - **E-Accessible Web tool for video contents (prototype)**
 - For playing video through a WEB server with selection of different VD levels according to user's preference
 - **Adaptive Video Description player (beta version)**



Project 1.2 & 1.3

Motivations

- VD is mainly a manual task
- Human “describers” watch the film and write a script that identifies key visual elements
- Carefully time the placement and length of the description to fit within nearby audio segments
- Up to 25:1 workload ratio
- CRTC requirements increase

- Blinds and vision impaired
 - USA: 10 millions ⁽¹⁾
 - Canada: 1 million ⁽²⁾
 - “Vision loss strikes 1 in 9 people over age 65, and this proportion climbs to 1 in 4 people over age 75. In Canada, the number of people living with blindness or vision impairments will increase by 52% by year 2026” ⁽³⁾

- (1) The American Federation of Blind: <http://www.afb.org>
- (2) Canadian National Institute of Blind: <http://www.cnib.ca>
- (3) Jacques Gresset, École d’Optométrie de UdeM



Projects 1.2 & 1.3

Main tasks

- E-inclusion - Phase 1
 - Development data
 - VD productions analysis
 - Screening sessions and interviews with end-users
 - Software development environment
 - Software prototype for VD production and adaptive VD player
 - Feedback from VD producers
- E-inclusion – Phase 2
 - Additional development & test data
 - Beta versions
 - Feedback from VD producers
 - Web accessible site setup for VD
 - Feedback from end-users

What the users want (Turner)



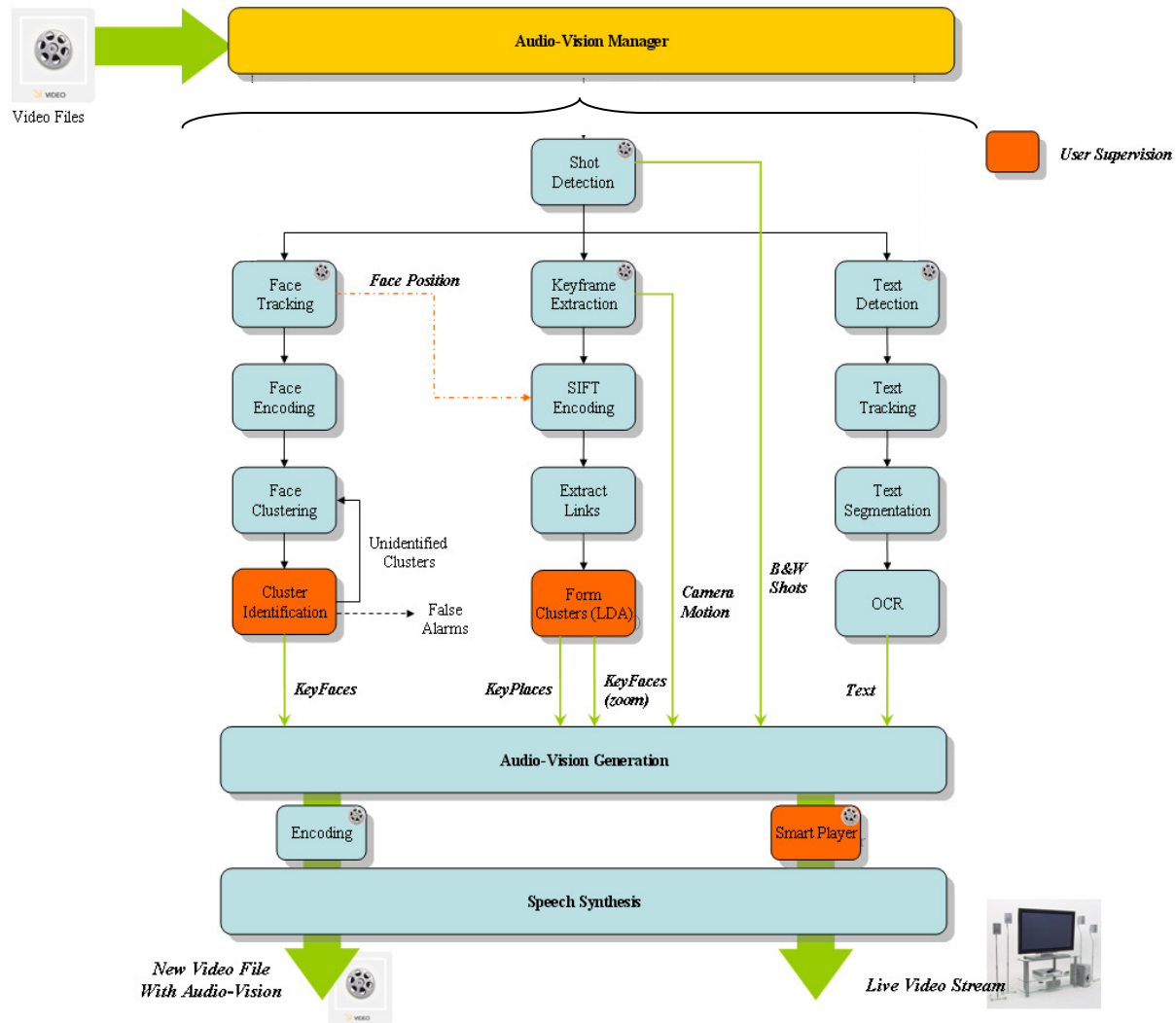
- Who is talking? (mainly at the beginning of the films)
- Who is present in the scene ?
- Emotional states (facial expressions)
- Scene location (day/night, indoor/outdoor, ...)
- Actor appearance (costumes, gesture)
- Scene changes (shot transition)
- Key-words (transition texts, credits and subtitles, ...)
- ...
- **Select the level of VD details...**
 - some end-users want a lot, others none at all !

Technical trade-off



- Very complex to make automatic
 - Highly semantic for some content
 - Highly uncontrolled environment
- We concentrate on
 - Shot transitions (cut, fade-in, fade-out)
 - Detection of main figures (key-faces)
 - Detection of main places (key-places)
 - Text spotting and OCR
 - Detection of camera motions
 - Gait & gesture

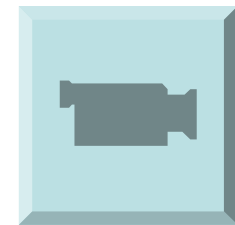
System integration



Video Description Output

```

(0)[wav] Deplacement vers la droite, vers le bas et eloignement
(205)[wav]
(285)[wav] Deplacement vers la gauche, vers le haut et rapprochement
(371)[wav] 6 sequences successives
(421)[wav] [421-1418] Amelie
(459)[wav] appartement d'Amelie
(600)[wav] Deplacement vers la droite et eloignement
(733)[wav]
(855)[wav]
(949)[wav]
(1005)[wav]
(1065)[wav]
(1173)[wav]
(1232)[wav] Deplacement vers la droite et rapprochement
(1324)[wav]
(1415)[wav] Deplacement vers le haut et rapprochement
(1705)[wav] Deplacement vers le haut
(1800)[wav]
(1837)[wav] Rapprochement
(2030)[wav] Deplacement vers la droite et rapprochement, appartement de l'epicier
(2382)[wav] que sur la villa
(2468)[wav] Deplacement vers la gauche et rapprochement
(2552)[wav] Deplacement vers le bas
(2724)[wav] Deplacement vers la gauche, vers le bas et rapprochement, pizzeria, Lucien
(3030)[wav] Deplacement vers la gauche et rapprochement
(3140)[wav] Rapprochement
(3308)[wav] Moved from frame 3382 Deplacement vers le bas
(3485)[wav]
(3512)[wav]
(3584)[wav] Cafe des deux moulins, Amelie,
(3698)[wav]
(3730)[wav]
(3754)[wav]
(3791)[wav]
(3851)[wav]
(7544)[wav]
(7563)[wav] Deplacement vers le bas
(7615)[wav] 2 sequences successives
(7660)[wav]
(7700)[wav]
(7795)[wav]
(7831)[wav]
(7862)[wav]
(7935)[wav]
(8004)[wav]
(8057)[wav]
(8100)[wav]
(8199)[wav]
(8314)[wav]
(8380)[wav]
(8514)[wav]
(8580)[wav]
(8544)[wav]
(2030)[wav] Deplacement vers la droite et rapprochement, appartement de l'epicier
(2382)[wav] que sur la villa
(2468)[wav] Deplacement vers la gauche et rapprochement
(2552)[wav] Deplacement vers le bas
(2724)[wav] Deplacement vers la gauche, vers le bas et rapprochement, pizzeria, Lucien
(3030)[wav] Deplacement vers la gauche et rapprochement
(3140)[wav] Rapprochement
(3308)[wav] Moved from frame 3382 Deplacement vers le bas
(3485)[wav]
(3512)[wav]
(3584)[wav] Cafe des deux moulins, Amelie,
(3698)[wav]
(3730)[wav]
(3754)[wav]
(3791)[wav]
(3851)[wav]
(7505)[wav] Deplacement vers la gauche, vers le haut et rapprochement
(7493)[wav] [No empty space found] Deplacement vers la droite
(7271)[wav] Rapprochement
(7304)[wav]
(7098)[wav] Deplacement vers la droite, vers le bas et rapprochement
(6961)[wav]
(6910)[wav]
(6257)[wav] Moved from frame 6817 Appartement du voisin peintre, Amelie, Gina, Le voisin peintre, Lu
(6614)[wav] Deplacement vers le haut et eloignement
(6381)[wav] Deplacement vers la gauche, vers le haut, et rapprochement
(5809)[wav] Deplacement vers le haut
(6178)[wav] Deplacement vers la gauche, vers le haut et eloignement,
(6114)[wav]
(5000)[wav] Deplacement vers la gauche et rapprochement
(4924)[wav]
(4810)[wav] Deplacement vers la droite
(4777)[wav]
(4723)[wav] Eloignement
(4674)[wav]
(4638)[wav]
(4565)[wav]
(4507)[wav]
(4464)[wav]
(4421)[wav]
(4354)[wav]
(4281)[wav]
(4250)[wav]
(4194)[wav]
(4105)[wav]
(4025)[wav] Deplacement vers la droite et rapprochement
(3974)[wav] Rapprochement
(3905)[wav] Deplacement vers la droite, vers le bas et eloignement
(11035)[wav]
(11032)[wav] 2 sequences successives
(10908)[wav]
(10797)[wav]
(10380)[wav] Deplacement vers la droite
(10350)[wav] 3 sequences successives
(10221)[wav] Deplacement vers la gauche et rapprochement
  
```



Audio Segmentation

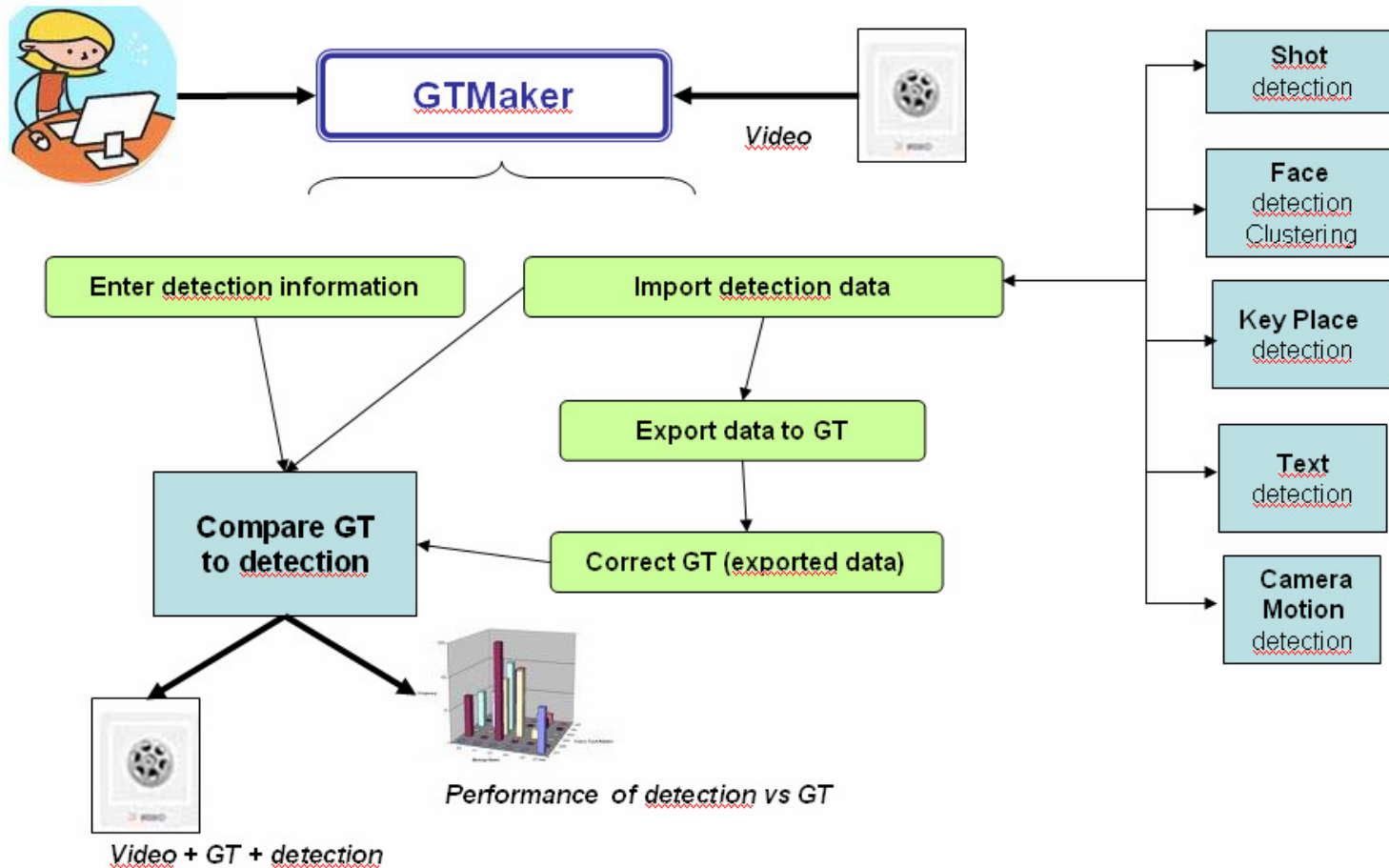
Text-to-Speech

Visual Content Detection

Time Tag

Text Description

Ground Truth Maker





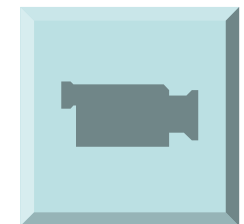
Ground Truth Maker

Beta version

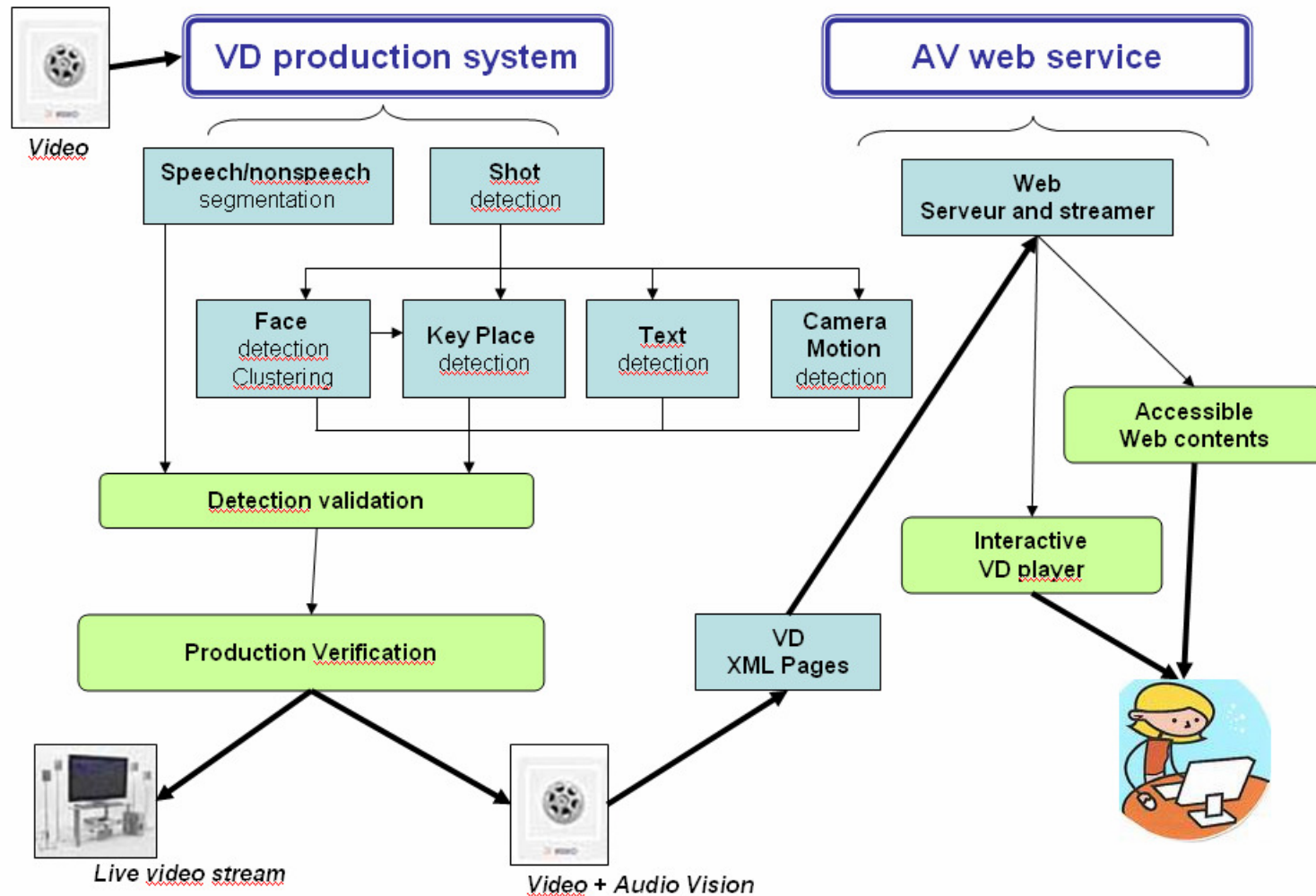
The screenshot displays the Ground Truth Maker software interface. The main window is titled "VV_Ep05_Ch2 - CRIM Ground Truth Maker (beta) v0.4.03038.0003". The interface is divided into several sections:

- Video Player:** Shows a video frame with two people. Green bounding boxes are drawn around their faces, labeled "G.T. 2 - 27" and "G.T. 1 - 29". Red bounding boxes are drawn around text overlays: "Amielle gérante" and "Marcelo chef cuisinier".
- Layer Selection:** A list of checkboxes for different ground truth layers, including G.T. Transition, Auto Transition, G.T. KeyFace, Auto KeyFace, G.T. Camera, Auto Camera, G.T. Text, Auto Text, G.T. Object, and Auto Object.
- Data Table:** A table with columns: Start, End, Category1, Category2, and Auto. It lists various ground truth events with their corresponding time ranges and categories.
- Camera Category:** A section for defining camera categories, showing a list with "Category1" and "Category2".
- Timeline:** A playback control bar at the bottom with a progress indicator and a "Stop" button.

	Start	End	Category1	Category2	Auto
25	5316	5438	combined		0
26	5439	5480	combined		0
27	5481	5599	combined		0
28	5600	5683	tilt up		0
29	6075	6221	combined		0
30	6222	6374	others		0
31	6375	6436	combined		0
32	6437	6480	combined		0
33	7857	7905	pan right		0
34	8078	8690	others		0
35	8691	8852	others		0
36	8854	9142	others		0
37	9144	9295	others		0
38	9773	9876	others		0
39	10788	11138	others		0
40	11908	12024	combined		0
41	12025	12197	combined		0
42	13341	13432	pan right		0
43	13730	13916	others		0
44	3375	3868	combined		0



Enhanced Talking Web Browser



Adaptive Video Description player *Prototype*

