

Text-to-Movie Authoring of Anatomy Lessons

Vaishnavi Ameya Murukutla, Elie Cattan, Olivier Palombi, Rémi Ronfard

▶ **To cite this version:**

Vaishnavi Ameya Murukutla, Elie Cattan, Olivier Palombi, Rémi Ronfard. Text-to-Movie Authoring of Anatomy Lessons. AIME 2019 - 17th conference on artificial intelligence in medicine, Jun 2019, Poznan, Poland. pp.304-308, 10.1007/978-3-030-21642-9_39 . hal-02143822

HAL Id: hal-02143822

https://inria.hal.science/hal-02143822

Submitted on 29 May 2019

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

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

Text-to-Movie Authoring of Anatomy Lessons

Vaishnavi Ameya Murukutla[1][0000−0001−9455−3391], Elie Cattan[2], Olivier Palombi[1], and Remi Ronfard[1][0000−0003−4830−5690]

[1] Univ. Grenoble Alpes, LJK, Inria, CNRS, 38000 Grenoble, France
[2] Anatoscope, 38330 Montbonnot-Saint-Martin, France

Abstract. With popular use of multimedia and 3D content in anatomy teaching there is a need for a simple yet comprehensive tool to create and edit pedagogical anatomy video lessons. In this paper we present an automated video authoring tool created for teachers. It takes text written in a novel domain specific language (DSL) called the Anatomy Storyboard Language (ASL) as input and translates it to real time 3D animation. Preliminary results demonstrates the ease of use and effectiveness of the tool for quickly drafting video lessons in realistic medical anatomy teaching scenarios.

1 Introduction

Anatomy is the cornerstone of medical education. Fundamental knowledge of the human body is essential for understanding other subjects in medical and para medical fields. Traditionally anatomy is taught using visual aids such as chalk board drawings and slide presentations. Previous studies have shown that 3D graphics and animation make anatomical learning more engaging [1] and effective [3, 6] but they suffer from a content creation bottleneck. If teachers choose to incorporate 3D animation in their lessons they either have to use the content already available to them or invest resources to create new content with the help of a graphic designer. In the first case content may not match the learning objectives of the class and the second case offers very little control to the teachers over the finished video. The solution to this would be to enable anatomy experts to generate their own 3D animations using innovative authoring tools.

Text-to-movie (or text-to-scene) authoring is a general class of methods that have been proposed for automatically generating 3D graphics and animation from text written by a domain expert. Recently, Hassani and Lee have proposed a review of text-to-movie research focusing on natural language [2]. While they provide a useful conceptual framework for our work, we chose to use a special-purpose authoring language, rather than natural language, in order to better address the needs of medical education.

In our authoring system (Fig. 1(a)) scripts for the lessons are written in a new formal language called the Anatomy Storyboard Language (ASL). The scripts are then parsed and translated part by part into hierarchical finite state machines (HFSM). Finally, state machines are executed in Unity 3D game engine to produce the desired animation at runtime.

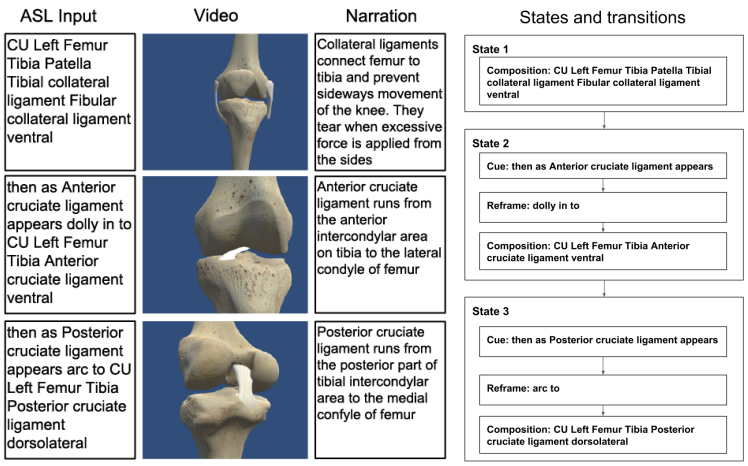


Fig. 1: Text-to-movie generation example with HFSM representation.

2 Anatomy Storyboard Language

It is a domain specific language that is both machine and human readable. The video to be produced is written as a set of unique sentences. Each sentence describes all the visual elements, camera actions and animations seen from the start of the recording till the camera stops. ASL is an anatomical extension of the Prose Storyboard Language [5] that was designed for annotating and directing movies. As each sentence is capable of generating a complete shot it must have all the information necessary to transition into the shot, build composition, direct camera movements and record changes in composition as subjects in the video perform actions. The complete And/Or graph of the ASL grammar is presented in Fig 2.

ASL is a context free language with terminals (anatomical entities and cinematographic terms) and non terminals (initial *composition* and subsequent *development*). The terminals are either generic terms used for camera movement or animation, or specific terms referring to the *subject* described in the shot such as anatomical parts and regions. The nomenclature of these specific terminals is derived from My Corporis Fabrica [4] (MyCF), an extensive ontology that describes structural and functional relations of different parts of human body. *Composition* is a description of all the elements that are seen in a particular frame. It needs to be comprehensive in detailing the size(Figure 3(b)), angle, plane of view, anatomical location or specification, profile and relative screen position of the subjects viewed. The subjects in our case are anatomical parts and regions as seen in the complete 3D male Zygote model for human body [3].

[3] https://www.zygote.com

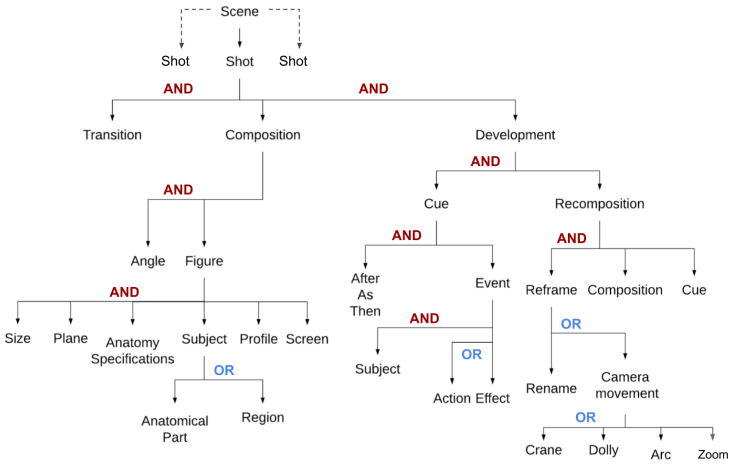
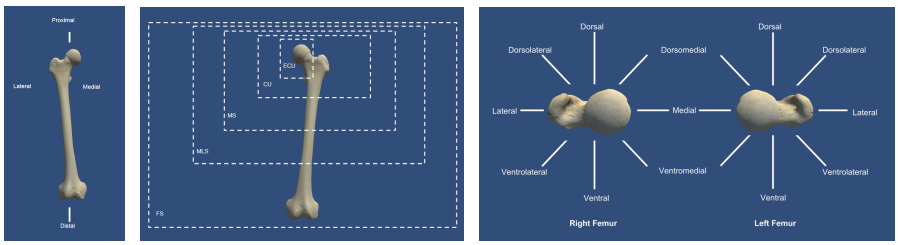


Fig. 2: And/Or Graph representation of the Anatomy Storyboard Language grammar. ASL scenes are made of shots containing an initial composition and one or more optional developments.

The most important descriptive elements that are essential in building the *composition* are *plane*, *anatomical specification*, *profile*. *Plane* refers to the hypothetical planes that divides the human body. In ASL they define the view in which we see the anatomical parts and direct camera position accordingly. If a plane is not mentioned in a composition then the system will automatically assign a plane in the vertical axis *(sagittal or frontal)* based on the *profile* information but if the desired composition is in horizontal axis *(transverse)* then it must be mentioned in composition.



(a) Specifications (b) Anatomy sizes (c) Profiles in transverse plane

Fig. 3: ASL specifications, sizes and profiles.

Anatomical specification (Fig 3(a)) further defines the part of anatomy most in focus in the current composition. *Profile* describes the orientation of the part

in relation to the camera. It is the side of the subject that is viewed by the camera. The anatomical profile of left and right femurs are shown in Fig 3(c). As the shot develops there will be changes in the composition. These changes can be due to *Actions* or *Effects* in *Cues*, or *Camera movements*, or both.

Sentences written in ASL are parsed via the Parsimonious[4] parser in the Python language. The parsed sentences are then translated to a hierarchical finite state machine (HFSM), with one state per composition or development. The HFSM is described in a XML format with specific tags to define each state of the machine. Finally, the HFSM is interpreted and executed to generate the desired animation. We now describe each of those steps separately.

3 From ASL to Animation

3.1 HFSM generation

The different elements written in ASL are organised into states and transitions of a HFSM. Particular tags are used to describe the state machine. Our Python based HFSM generator creates a *scenario* tag for each complete ASL sentence. A *scenario* tag contains a list of *states* and *transitions*. Each *state* is given a unique *name* and an *anatomy* list of parts present in the current composition. A *state* also describes a *camera* with several tags that characterise its positioning such as *orientation*, *angle_up*, *angle_side*, *up* . In particular, a *lookat* tag lists the objects the camera should look at. A *transition* is a change from a *start* state to an *end* state. Currently the transitions are executed automatically between two consecutive states with a preset time delay that is specified in the *delay* tag.

Actions in ASL are translated into animations in the HFSM. An additional *animation* tag is added in *state* to trigger animations of anatomical elements (e.g. a knee flexion). In the current state of the application, animations are pre-made and cyclical. This is done to avoid editing glitches that could arise if the body position at the end of animation in one state does not match the body position in the next state.

Some descriptive terms of the ASL need to be converted to numerical values in HFSMs. For example, for the lessons written in this paper we specified that the ASL term *high angle* will be translated to a 45 degrees bird's eye view. This numerical value of 45 is defined in an animation style sheet along with other global values that change the camera position and total run-time of the video. This style sheet can be edited by the teachers depending on their preferences thereby giving them more nuanced control over the video making process.

3.2 Animation generation

We developed an application using the Unity 3D game engine to generate the desired animation at runtime from the HFSM obtained from the ASL script. The application is thus an interpreter, from a specific XML format to 3D videos

[4] https://github.com/erikrose/parsimonious

of anatomy. The description of the *camera* in the HFSM is given by a view plane (frontal, sagittal or transverse), an object or a group of object to look at and a up vector to orient the rotation of the camera. With these pieces of information, the application computes the bounding boxes of the objects to look at and deduces the position and orientation to reach. The camera then moves from its previous position to the new one according to the other parameters translated from the ASL (e.g. type and speed of camera movement). If an *animation* tag is present, it executes an animation from the database that is registered under the same name.

4 Results and future work

We used our text to movie authoring system to create short videos based on scripts written by two anatomy professors. The teachers were given a brief introduction to ASL using the And/Or chart. Examples of compositions in ASL and their corresponding frames in Unity player were shown to get them familiarised with the system and ASL grammar. They initially started writing very short scripts with initial composition and one development. Progress was made one composition at a time during which they tested different animations and decided on the best viewing positions. The most liked features of the system are that it allows the user to build the video state by state and has immediate visualisation of the video made so far which facilitates easier editing. After some practice the teachers were able to write three lessons on the knee joint and one lesson on the forearm.

In future work, we would like to extend our approach to non-linear content generation by taking into account user-triggered transitions between states. This would make our approach applicable to mixed reality. Another promising direction for future research would be to generate ASL scripts directly from audio narrations, using a combination of speech processing and natural language understanding.

References

1. Estai, M., Bunt, S.: Best teaching practices in anatomy education: A critical review. Annals of Anatomy - Anatomischer Anzeiger **208**, 151–157 (nov 2016)
2. Hassani, K., Lee, W.S.: Visualizing natural language descriptions: A survey. ACM Comput. Surv. **49**(1) (Jun 2016)
3. Hoyek, N., Collet, C., Rienzo, F.D., Almeida, M.D., Guillot, A.: Effectiveness of three-dimensional digital animation in teaching human anatomy in an authentic classroom context. Anatomical Sciences Education **7**(6), 430–437 (mar 2014)
4. Palombi, O., Ulliana, F., Favier, V., Léon, J.C., Rousset, M.C.: My corporis fabrica: an ontology-based tool for reasoning and querying on complex anatomical models. Journal of Biomedical Semantics **5**(1), 20 (May 2014)
5. Ronfard, R., Gandhi, V., Boiron, L.: The prose storyboard language: A tool for annotating and directing movies. CoRR **abs/1508.07593** (2015)
6. Vernon, T., Peckham, D.: The benefits of 3d modelling and animation in medical teaching. Journal of Audiovisual Media in Medicine **25**(4), 142–148 (2002)