



HTML5 Case Study 6:

3Dactyl: Using WebGL to Represent Human Movement in 3D

Document details

Author :	Stephen Gray
Date:	21 May 2012
Version:	V1.0
Rights	This work has been published under a Creative Commons attribution-sharealike 2.0 licence.
About	This case study is one of a series of HTML5 case studies funded by the JISC which provide examples of development work in use of HTML5 to support a range of scholarly activities.

Acknowledgements

UKOLN is funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from the JISC and the European Union. UKOLN also receives support from the University of Bath where it is based.

Contents

1 About This Case Study	1
Target Audience	1
What Is Covered	1
What Is Not Covered	1
2 Use Case	2
Background	2
3 Solution	3
Motion capture	3
Motion Re-targeting	3
Re-presenting Performance as X3D	3
4 Challenges	5
Quality of motion capture	5
Use of surrogate CGI avatars vs. 3D 'scanning'	5
Delivery	5
5 Things Done Differently / Lessons Learnt	6
6 Conclusions	6
3D as Evidence of Research	6
3D as an Archival Document	6
3D as Student 'Sketchbook'	6
Plans for Future Development	6
Appendix 1: HTML5 Technologies Used	7

1 About This Case Study

This case study covers the development of 3Dactyl, a hardware and software configuration, which is intended to record and represent the physical movements of an individual online in three dimensions, for scholarly research purposes. Resulting 3D scenes (as an XML document) are embeddable within a standard Web page or VLE. Examples of such 3D footage might be various forms of performance art, e.g. dance, drama or even sport where the performance of play strokes can be carefully analysed. Within the same constraints of space, surgical or therapeutic procedures would be another feasible use. When such scenes are viewed on future versions of browsers, they will not, typically, require special plug-ins to use the 3D footage interactively.

The 3Dactyl Project is still under development, the aim is to develop a system which is reproducible, configurable and usable by the non-technical specialist at minimal cost. Future versions will aim to automate much of the workflow, which for now has to be carried out manually.

Research groups from medicine, sports science and archaeology are beginning to explore the presentation of research data as online 3D visualisations. New precedent projects demonstrate a clear need for spatially accurate information, presented over time in a simple and accessible way.

The particular focus for the current project has been the capture and representation of human motion within arts research. For example, the tracking of human motion in the context of dramatic action, stage behaviour, etc. To achieve this, we have worked closely with the University of Bristol's Department of Drama: Theatre, Film, Television¹.

Target Audience

This case study will be of interest to creative arts researcher-practitioners within Higher Education or similar research institutions. Although the configuration and operation of the 3Dactyl system currently does require a degree of technical engagement it is being specifically developed by the 3Dactyl team for use by arts researchers who are not IT specialists.

It is hoped that those responsible for supporting performance-related research activities such as: technical theatre staff, assessors of practice-as-research, keepers of performance archives and e-learning staff responsible for arts faculty VLE pages will also find this case study useful and may wish to reproduce the workflows described below.

What Is Covered

The case study divides the development of 3Dactyl into three phases:

1. Motion capture
2. Re-targeting motion data to a standardised digital avatar
3. Representing the avatar with real-world motion applied in an interactive 3D form, via a browser

The hardware, software and workflow steps for each phase are described in this case study.

What Is Not Covered

The current incarnation of 3Dactyl requires a 3D avatar model to be used in order to 'carry' the 3D data which has been captured. Many suitable 3D models are available on the web under Creative Commons licences. Use of custom avatars is possible and has exciting creative possibilities; but CGI modelling, rigging and skinning is beyond the scope of this document.

¹ University of Bristol: Department of Drama: Theatre, Film, Television, <http://www.bristol.ac.uk/drama/>

2 Use Case

In addition to enriching our cultural heritage sector, performance, as research, underpins the scholarly record and is commonly used, reused and reinterpreted by subsequent researcher-practitioners as the basis for new works. The target audience for this system is the undergraduate through to a post-doctoral researcher studying in a performance-related discipline (e.g. theatre, live art or dance).

Locally, the system is intended to be used by undergraduates studying performance within the University of Bristol's Department of Drama to build 3D 'sketchbooks'. 3Dactyl is presented as a solution which will permit 3D recordings to be made, studied and ultimately archived for research purposes.

Students, researchers and keepers of performance study collections recognise both the desirability and the considerable challenge of 'preserving' live performance. Often a single video recording is used to represent a work which may have taken months or even years to develop. As 'by-products' of the creative process, these videos may be of very poor quality. Problems are associated with using any single method of documentation, but video has an inherent limitation: it is essentially a 2D technique attempting to describe actions in 3D space. For disciplines where correct execution is important, such as dance, established visual recording methods frequently produce documents which are entirely unfit for research or teaching purposes.

42 UK Higher Education institutions (HEIs) and similar institutions put forward research for assessment under the Drama, Dance and Performing Arts heading of the 2008 RAE. The University of Bristol Department of Drama: Theatre, Film, Television was ranked 6th among them. The University is also home to the Theatre Collection, a special study collection and museum which holds the second largest performance-related archive in the UK, after the Victoria and Albert Museum.

After conducting several successful collaborative projects with the Department of Drama the 3Dactyl team recognised the need for an inexpensive and easy-to-use system which could be deployed to record performance in an interactive 3D format.

The system should be unobtrusive (for example, not require special markers to be worn during recording) and sufficiently accurate, as well as be able to produce documents in a format which can be accessed via a browser without the need for dedicated plug-ins. The solution will allow students and researchers of performance to interact with 3D representations in real time and to examine the event from any conceivable perspective, for example, from behind, above or at very close range. This need ruled out the use of domestic 3D TV technologies, which record and display stereoscopic rather than truly 3D representations, in much the same way as conventional cinematic projections differ from 3D vision performances.

If possible, documents produced by the system should be in an open source format in order to meet the collection policies of archives such as Bristol's Theatre Collection Museum² which is the custodian of much of the UK's performance documentation.

Background

Tools which facilitate online 3D representations have been around for some time. Introduced in the early 1990s, VRML (Virtual Reality Mark-up Language) was limited by insufficient bandwidths, processor speeds and the need to employ browser plug-ins which proved difficult to configure. Despite these issues, VRML ultimately proved to be a successful, open technology and spawned the newer X3D format. The X3D ISO standard, used by the 3Dactyl system, offers the ability to encode a 3D scene using an XML dialect. Supporters of X3D are currently striving to make it the de facto standard for interactive 3D Web content.

X3D, coupled with advances in WebGL (Web-based Graphics Library) and intermediary technologies mean many of today's browsers are already capable of displaying interactive 3D data without the need for plug-ins, although this functionality is rarely used.

The online re-presentation of human motion is only possible if that motion has been accurately captured, to achieve this, the current project makes use of the Microsoft Kinect, an inexpensive

² University of Bristol Theatre Collection, <http://www.bris.ac.uk/theatrecollection/>

peripheral, primarily intended as a controller for the Xbox 360. The release by Microsoft of the Kinect SDK³ has facilitated the development of many Mocap (motion capture) applications which would, only a short time ago have been financially prohibitive for the majority of academic researchers or even modestly funded research groups.

3 Solution

Motion capture

The project uses the (sub-£100) Kinect, a device intended primarily as a controller for the Xbox 360 game console. Motion capture also requires a PC, with OpenNI framework, NITE middleware and Brekel Kinect Drivers (all either open source or freely available). In the current version of 3Dactyl, the Brekel application is used as the central interface for motion capture. After launching the application, users stand in front of the Kinect device and ensure their entire body is visible. The human form is recognised automatically. Note however that recognition fails if a user's face is not visible to the device or if the user stands too close to large objects. In order to 'lock on' to a skeleton, the user must adopt a standard 'Psi-Pose' (see Figure 1).



Figure 1. The psi-pose

The Brekel application allows either visible light or infra-red light to be used for tracking and this option for users can make a difference to the quality of the outcome, depending on the lighting conditions of a room. Loose clothing or other people in the device's view can also make calibration fail (although tracking of multiple skeletons is possible and will shortly be implemented). When tracking starts an overlay of joints is visible in the viewport. The user (or an operator) chooses where the mocap data should be saved to and starts the capture process. The system begins to write a Biovision Hierarchy (BVH) character animation file, which contains the motion data, i.e., successive captures of the human movements which are numbered incrementally. Limitations to motion capture include a restricted 'stage' area within which the subject can move, as well as the requirement that the subject's *entire* body remains within those limited boundaries throughout the entire take.

The size of this area depends greatly on local lighting conditions, but 10 square feet of floor space can be expected. Another limitation is the inability to track fine-grained movements (such as finger movements or facial expressions), though this aspect is improving with successive generations of tracking software.

Motion Re-targeting

Once motion is captured in .BHV format it can be applied to a surrogate avatar. The avatar does not have to have the same proportions as the performer.

Nor does it have to be humanoid. Several 3D modelling packages allow the re-targeting of .BHV data onto an existing CGI model. We typically elect to do this via the open source modelling software, Blender. The same process is possible in packages such as Maya or 3DS Max. This stage is fairly standard and many guides exist which cover the application of .BHV data within specific modelling packages. One advantage of using Blender is the package's native ability to export a scene as X3D format, though both MAYA and 3DS Max can export as X3D via plug-ins (RawKee and BSContact, respectively).

Re-presenting Performance as X3D

3Dactyl relies upon three key technologies in order to deliver 3D content via a Web browser: X3D, 3DOM and WebGL.

³ Microsoft Kinect SDK for Developers, <http://kinectforwindows.org/>

X3D⁴ is the ISO standard XML-based file format for representing 3D computer graphics. X3D supports many of the same features as its predecessor, VRML, including: several different options for navigating a scene, the ability to loop animated elements and clickable (onclick) 3D objects, used to initiate actions. We selected X3D for use within 3Dactyl as the format is already becoming integrated into the HTML5 standard. The HTML5 specification⁵ already references X3D for declarative 3D scenes. However, a specific integration model is not suggested.

```

<!DOCTYPE html>
<html>
  <head>
    <meta http-equiv='Content-Type' content='text/html; charset=utf-8'></meta>
    <title>X3DOM example
    </title>
    <link rel='stylesheet' type='text/css' href='http://www.x3dom.org/x3dom/src/x3dom.css'></link>
  </head>
  <body>
    <h1>X3DOM example
    </h1>
    <p>
      <x3d id='someUniqueId' showStat='false' showLog='false' x='0px' y='0px' width='400px' height='400px'>
        <scene DEF='scene'>
          <worldInfo info='"X3D sample created using the 3Dactyl system"' title='climb1'></worldInfo>
          <navigationInfo type='"EXAMINE" "ANY"'></navigationInfo>
        </scene>
      </x3d>
    </p>
  </body>
</html>

```

Figure 2. Code snippet from a finalised X3D file.

X3DOM⁶ is the DOM-based HTML5/X3D integration model we use, X3DOM integrates X3D data with HTML5 using only WebGL and JavaScript. This is achieved by directly mapping live DOM elements to an X3D model in a way very similar to the way SVG is used with 2D graphics.

WebGL⁷ displays the X3D scenes within the browser. WebGL uses the canvas element and provides a 3D graphics API. Many browsers which support HTML5 also have support for WebGL (see Table 1).

WebGL browser support	
Mozilla Firefox	Support enabled by default since version 4.0
Google Chrome	Support enabled by default since version 9
Safari	Support disabled by default since 5.1
Opera	Only supported in development build 11.50
Internet Explorer	Works via IEWebGL plug-in. No official plans to support WebGL without plug-ins.

Table 1: WebGL browser support.

Encoded as HTML5, X3D scenes can either be associated with other, non-3D content, or remain independent.

⁴ Web3D Consortium: X3D and Related Specifications, <http://www.web3d.org/x3d/specifications/>

⁵ HTML5: A vocabulary and associated APIs for HTML and XHTML, <http://dev.w3.org/html5/spec/>

⁶ X3DOM, <http://www.x3dom.org/>

⁷ WebGL - OpenGL ES 2.0 for the Web, <http://www.khronos.org/webgl/>

4 Challenges

Quality of motion capture

Fine-grained motion is still difficult to record reliably using the Kinect. This problem is by no means insignificant but should be balanced against the low cost of the unit. The Kinect is approximately 1000 times cheaper than a standard movie-grade motion capture hardware set-up but, under favourable conditions, can give comparable results. Since Microsoft released the Kinect SDK into the public realm, there has been rapid development of 'next-generation' motion capture applications which already demonstrate an improved ability to capture fine-grained motion data.

Use of surrogate CGI avatars vs. 3D 'scanning'

Performance is covered by a heterogeneous group of disciplines and, while there is some commonality, each views the action of documenting practice differently. To summarise the feedback the team received: researchers from theatrical disciplines viewed the 3Dactyl system as a opportunity to publish theatrical events (i.e. to an audience) while dance researchers saw the potential of 3D recording as a way to refine their practice by detecting and then correcting errors in technique. Other disciplines (e.g. live art) fell somewhere between these two camps.

The ideal system in the view of theatrical disciplines would require less accuracy in terms of motion capture but more CGI modelling, inclusion of scenographic details (such as a virtual 'stage' and props) and the interaction of multiple performances. For dance researchers, the ideal system would instead require accurate motion capture but less 3D modelling; indeed, in order to compare technique a process of standardisation was requested, which would allow two performances to be compared and contrasted simultaneously, regardless of differences such as body shape, costume or gender.

The question was therefore one of driving 3D avatars vs. the 3D 'scanning' of scenes, both were possible within the scope of the project but limited resources meant it would not be possible to investigate both avenues simultaneously. It was decided that the 3D avatar model should be investigated first and could later be used as the basis for more 'publishable' forms of 3D recording, possibly constructed using the technique of photogrammetry⁸.

Delivery

Over the course of 2011, several security issues have emerged which, according to Context Information Security, are inherent in the WebGL standard. The Context report⁹ claimed that security vulnerabilities were present in both the Chrome and Mozilla Firefox WebGL implementations. This meant that remote code could be employed to access local graphics software and hardware.

Mark Shaver, Vice President of Technical Strategy at Mozilla, stated that his organisation was working to address issues in the WebGL specification and Firefox's implementation of it¹⁰. Shaver went on to say that the Web needs a clear 3D standard and any such technology would inevitably experience teething problems.

Aside from IE's lack of support for WebGL (for example, no use of plug-ins), the security problems associated with WebGL have had little impact on the project as other major browsers have not withdrawn support for the standard. It is widely expected, given the desirability of presenting 3D data online without the need for plug-ins, that even if WebGL were to be withdrawn, another similar standard would quickly replace it.

⁸ *What is Photogrammetry?*, <http://www.photogrammetry.com/>

⁹ *WebGL - A New Dimension for Browser Exploitation*. Report, Context Information Security. Forshaw, J. May 2011., <http://www.contextis.com/research/blog/webgl/>

¹⁰ *Mozilla rejects Microsoft criticism of WebGL: New capabilities bring new risks*. *The Inquirer*. 21 June 2011, Wilson, D., <http://www.theinquirer.net/inquirer/news/2080571/mozilla-rejects-microsoft-criticism-webgl>

5 Things Done Differently / Lessons Learnt

User needs analysis is vital to successful implantation of innovative technologies. We judged our initial solutions to be unsuccessful, despite the fact that the results seemed to impress non-arts specialists. Feedback told us that solutions were too complex and (given the independent nature of much arts research) that only when users could understand and replicate the system and associated workflows from capture to online delivery would they view our solution as a serious research tool.

In retrospect, we should have maintained a far closer relationship with our user group members. Intermittent contact did not allow us to develop features which they believe would have been useful (such as the ability to annotate 3D scenes). Greater contact with the user group would have speeded development but, as users were unpaid volunteers, this proved difficult to achieve.

The wider academic application of the technologies discussed here did not occur to the team until late on in the development process. Again in hindsight, we should have put together a far wider user group. Representatives from the University of Bristol's Faculty of Veterinary and Medical Sciences and would have been particularly welcome additions to the project.

The long-term preservation of 3D documentation within archival repositories is something which interests the team greatly, and we are concerned about the long-term viability of such rich digital documents. On reflection, we should have engaged more closely with parallel projects which aim to address these issues such as the JISC-funded POCOS Project.

6 Conclusions

3D as Evidence of Research

As the Research Excellence Framework (REF) 2014 approaches, new ways are growing in importance which address the failings that are only identified during the periods of assessment by disciplines which deal with 'liveness' or dynamic animal behaviour. Performance, in the same way as any other academic discipline, has to demonstrate impact and excellence. A key benefit offered by our system to potentially adoptive departments is the ability to record and then represent works carried out by research staff during the course of their observations. External assessors will require no special plug-ins nor complex instructions to view 3D recordings of live work.

The representation of information in 3 dimensions is not always appropriate. However, our work has shown that when 3-dimensional documentation of performance is presented alongside other information (such as a catalogue records, videos or photographs) the multiple approaches can complement one another and are of great interest to our user community. The result is a hybridised and augmented document which is generally regarded as far richer than the sum of the parts.

3D as an Archival Document

It is our intention soon to publish a number of significant live works online in three dimensions. Feedback from academics has indicated that, once available, such a resource may be unique and is certainly of immediate and direct scholarly value. Subject to funding, we then intend to expand the collection to include more works.

3D as Student 'Sketchbook'

The building and embedding of student's own 3D 'sketchbooks' has already been the subject of a limited pilot and on the whole works well from a technical perspective. Building such activity into existing curricula is likely to take far longer to achieve and depends on the activity having proven academic benefit. Work is currently underway to identify a programme (such as an undergraduate course) willing to engage in a more in-depth trial.

Plans for Future Development

The next stage of the project will involve development of a Kinect-to-HTML software application, to be used by the arts researcher. This application will automatically retarget motion data to a

standard avatar and then generate required XML within an HTML document. The researcher then need only share this Web page via current conventional means.

Following useful feedback we intend to remove the explicit reference to a script in all 3D content files we produce. So:

```
<x3d></x3d>
<script type='text/javascript' src='http://www.x3dom.org/x3dom/src/x3dom.js'></script>
```

becomes:

```
<section itemtype="http://to-be-confirmed" itemscope>
<x3d>... </x3d>
</section>
```

thereby allowing the CMS or VLE serving the page to add the script manually or via a client-side JavaScript. These measures guard against the possible relocation of explicitly referenced script.

Future developmental stages will explore the use of multiple webcams as capture devices, replacing the Kinect. This would allow data to be reconstructed as photorealistic 3D representations via the process of photogrammetry (and so addressing the preferred solution of theatrical disciplines mentioned above). As the resulting files would be of considerable size, questions around bandwidth remain to be addressed. Off-line use of these highly detailed 3D documents remains an option.

The use of photogrammetry also offers an intriguing possibility in terms of archival recordings of performances which were filmed from multiple camera angles. 3D data could theoretically be derived and a 3D 'version' of an historical performance be created. We intend to investigate the scholarly benefits of this 'retrospective' motion capture process in later development phases.

Appendix 1: HTML5 Technologies Used

WebGL

WebGL is a cross-platform, royalty-free Web standard for a low-level 3D graphics API based on OpenGL ES 2.0, exposed through the HTML5 Canvas element as Document Object Model interfaces. Developers familiar with OpenGL ES 2.0 will recognise WebGL as a Shader-based API using GLSL, with constructs that are semantically similar to those of the underlying OpenGL ES 2.0 API. It stays very close to the OpenGL ES 2.0 specification, with some concessions made for what developers expect from memory-managed languages such as JavaScript.¹¹

Canvas element

The canvas element provides scripts with a resolution-dependent bitmap canvas, which can be used for rendering graphs, game graphics, or other visual images on the fly. In interactive visual media, if scripting is enabled for the `canvas` element, and if support for `canvas` elements has been enabled, the `canvas` element represents embedded content consisting of a dynamically created image.¹²

Document Object Model (DOM)

The Document Object Model is a platform- and language-neutral interface that will allow programs and scripts to access and update the content, structure and style of documents dynamically. The document can be further processed and the results of that processing can be re-incorporated into the presented page.¹³

¹¹ WebGL - OpenGL ES 2.0 for the Web, <http://www.khronos.org/webgl>

¹² 4.8.11 The canvas element: HTML Standard, <http://www.whatwg.org/specs/web-apps/current-work/multipage/the-canvas-element.html>

¹³ W3C Document Object Model (DOM), <http://www.w3.org/DOM>