VIOS: Vision Interactive Operating System

1. Project Summary

The goal of this project is to change the television viewing experience into an immersive and interactive activity that is tailored to the viewer's interest in the least invasive manner. The project focuses on in-video product embedding, tagging and linking for less invasive continuous advertisement and more effective user interaction so that a user can click on many objects *in* the TV show to get more information or be directed to the product purchase site. The project can expand by Integrating physical aspects such as mood lighting, so that the entire room is activated when an appropriate context-aware video signal is received (e.g. When the Yankee's hit a home run, the sports bar lights up).

Our solution will focus on the authoring and runtime system for an *Interactive Advertisement* application, the user will be able to purchase products online directly from the TV show. Our efforts will be on digital video authoring tools for object definition, tracking, embedding, tagging and linking to related product content. This interaction will be supported by on-screen graphics and interactive control. The *Immersive Ambience* application will deliver a rich viewing experience with autonomous solutions like appliances control, mood lighting and integrating it all on one box. In this way we can merge the virtual and the real world.

This new project is done in close collaboration with Comcast Cable's Office of the CTO.

2. Challenge Definition

TV watching is a passive activity and will remain a passive activity. However, the traditional approach of interrupting the show for advertisements is both invasive and inefficient as the user has limited capabilities in conveying their preferences for products and has no way to directly purchase the product during the commercial break. The goal of our project is to enable the viewer to purchase almost anything within the TV show (i.e. the user selects the product and not vice versa). The products embedded in the show do not interrupt the show as they are continuous discreet advertisements.

The video stream is tagged with the available products in the current frame and activation of the product label will transfer the user to the embedded link for more information. We thus aim to develop the authoring tools for product label marking, embedding, tagging and linking within any video show, at design time. At runtime, we will demonstrate both product tags in the video stream. Product tags will facilitate the above buyer activity. Context tags (e.g. the main actors finally get married in the episode or when the Steelers score a touchdown) which can also be embedded in the video

stream will trigger an immersive ambient experience. This solution will be useful with home-entertainment systems, restaurant and bar entertainment systems and also in educational and electronic advertisements in kiosks.

Our approach will be to develop a digital content tool-chain for semi-automated authoring of tagged digital content which will be HTML5 video compliant so that Packet Identifiers (PIDs) will encode the spatio-temporal placement of the active embedded products in the broadcast video stream. We will demonstrate the Interactive Advertisement aspect of the project with real TV shows.

Similar solutions, provided by companies such as Abid Technologies, are not completely automated and also do not provide for an immersive experience. Our system, VIOS, on the other hand, will provide an integrated cyber-physical interactive solution. Furthermore, our approach will be developed in collaboration with Michael Cook, from the Comcast Cable Corporation's CTO's Office. To summarize, we aim to provide an efficient solution to build an interactive TV around an actuated environment to enhance the user experience.

3. Proposed Solution

The proposed solution will involve:
Digital Signal Processing in the form of computer vision algorithms on the video stream;

Video Content Processing for embedding, tagging and linking metadata within the HTML5 supported video streams; and

Cyber-Physical Processing for runtime tag identification.

A. Vision Processing: Object detection is an extremely difficult challenge especially in real time video streams. There are a large number of classifiers present to do the same but each of them has different hit rates and false positives based on the number of training samples given. There is no one particular classifier which will work the best and hence it is necessary to combine them to improve results otherwise we may end up with a large number of false positives and very bad hit rates. Our solution aims to be robust and take care of many of these problems.

Much of our solution will be developed on open-source libraries such as OpenCV. We plan to implement the detection of faces on video streams using a hybrid classifier of fisher, eigen and LBP(Local Binary Pattern) classifier. Detection of various objects will be done using a Haar Cascade Classifier. Each of these classifiers can take upto weeks for training, so that the false positive rate goes down to the required level. In this way we can ensure that the object or face we want to detect based on an input for the TV will happen correctly throughout the course of the entire video.

B. Video Content Processing: This includes the embedding, tagging and linking of meta data within the video stream as Program Identifiers (PIDs). For a given video stream, the PIDs will be processed to have on-screen and cyber-physical responses. To create an immersive experience we can adjust the lighting according the context in the video stream . When the consumer points to a product on the screen, the tag pops-up on screen with a logo and link to more information. If the product being pointed is available in different color options the lighting can cycle through those colors continuously. Home automation can also be included so as the consumer can control the lighting and thermostat from the touch of TV remote itself.

C. Cyber-Physical Processing: While TVs today come with ambient background lighting, they do so by determining the average color in the current frame. We propose to take this concept to a more immersive level by allowing the video stream to embed context PIDs that have more relevance to the events within the TV show or that would appeal to the viewer. These can be based on the status of a sports game, a cliffhanger scene in a movie, the outcome of a Presidential debate, etc. Such semantically rich context will then trigger lights in the room.

4. Performance Measures

The Project would assume its final form, when we are able to:

Effectively track objects or faces in a video stream. While the embedding and tagging with metadata is done at design time since object identification and tracking in real-time for stream video is processor intensive and not necessary. At runtime, PIDs will identify the spatial and temporal location of the tags and highlight them with subtle object outlines.

Have the user view the tags by recognizing patterns and must be able to interact with the tags to trigger an on-screen or off-screen response. We will measure the errors due to mistaken tagging, errors in the remote-controller based or the voice recognition based tag identification by the user, the response time in opening the linked content and making a “purchase”.

Quantitative metrics which would be used for the measurements include:

Accuracy: The accuracy of tracking faces or objects. The accuracy with which tags placed are triggered.

Response time for tracking all tagged objects in a video stream

Cost: The total cost of the system.

5. Timeline & Milestones

Milestone	Deliverable	Due Date

Track faces of selected characters in primetime shows	• Recognize and Tag Characters of prime time show The Big Bang Theory • Repeat the tagging across multiple seasons to show consistency and robustness	**11/05/12**
Tracking along with detection	• Include Face tracking and alignment along with recognition • Improve detection rate and reduce false positive rate • Improve time of detection so that its close to real time	**11/12/12**
Detect frequently occurring objects in commercials and primetime shows	• Train LBP and Haar based Cascade classifiers to detect objects. • Tune the classifier to improve hit rate and reduce false positive rate • Include tracking	**11/19/12**
Develop Authoring tools in compatible with intel processor	• Integrate face recognition and object detection modules • Develop a common library and authoring tool	**11/26/12**
Ambient Environment Control	• Control Lighting based on digital content • Control Projector based on digital content (if time permits)	**12/2/12**

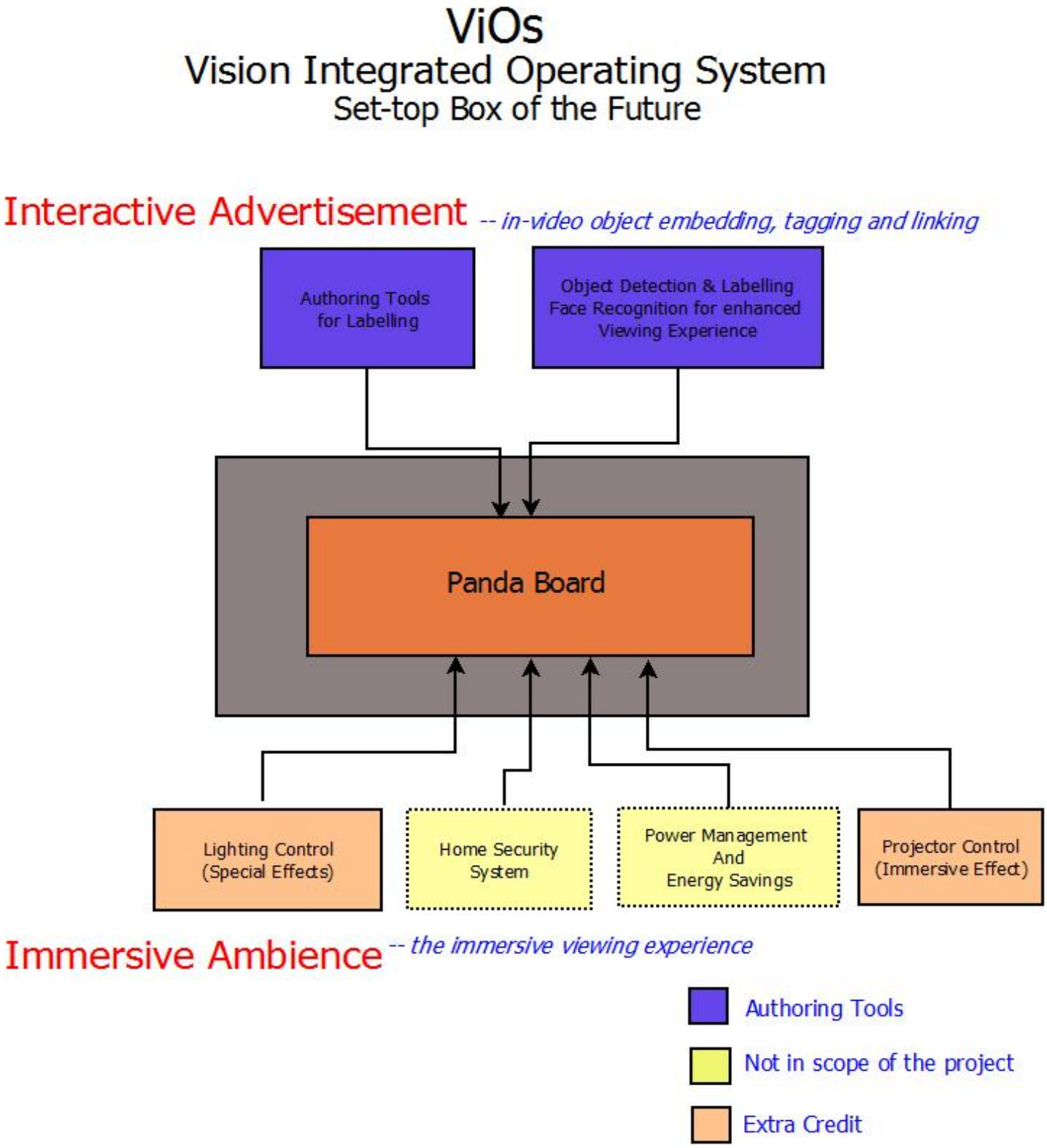

Fig 1. Overview Diagram

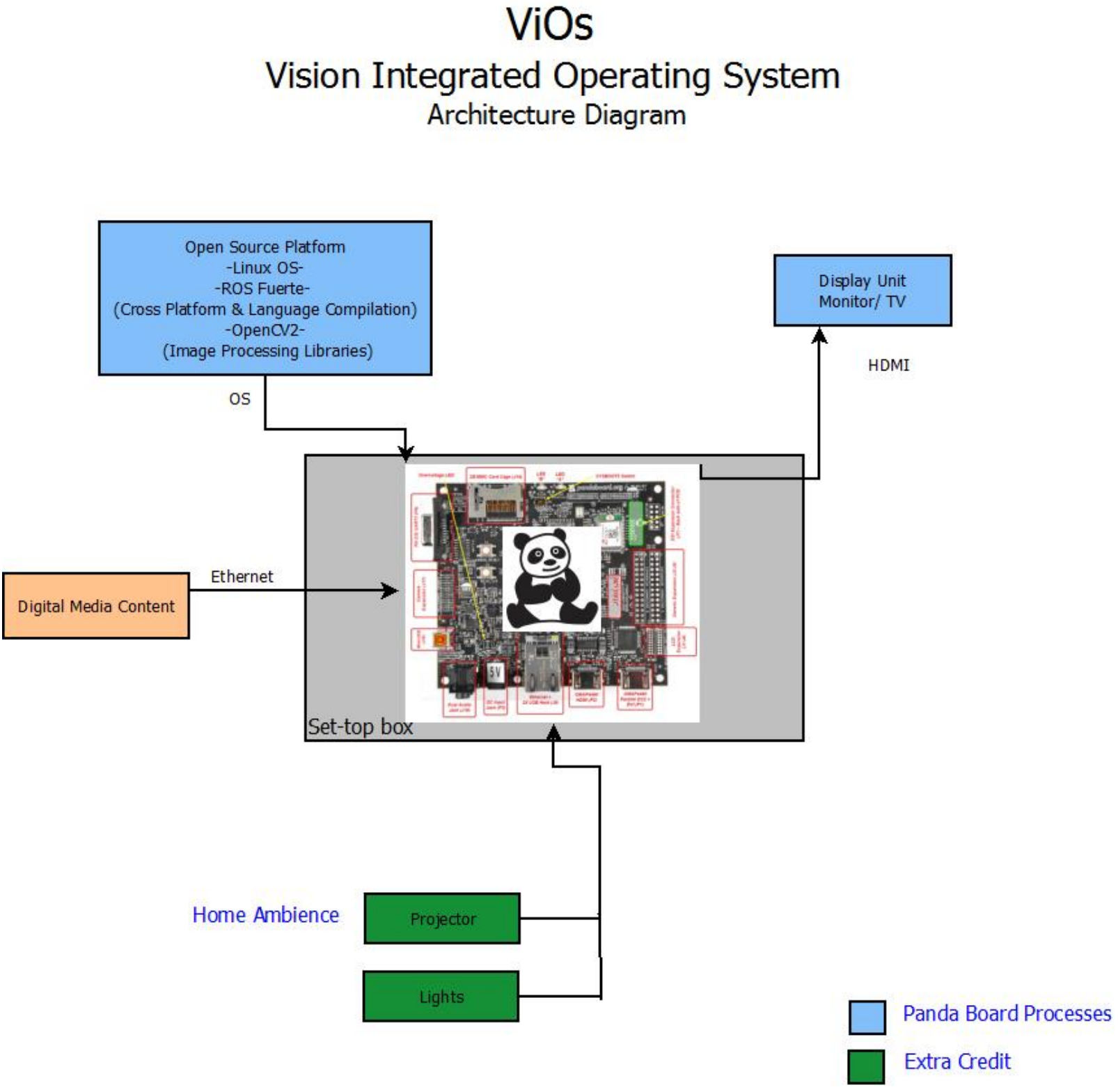

Fig 2. Architecture Diagram