Baseline Avatar Face Detection using an Extended Set of Haar-like Features

Darryl D'Souza, Roman V. Yampolskiy

Computer Engineering and Computer Science
University of Louisville
Louisville, Kentucky, USA 40292
darryl.dsouza@louisville.edu, roman.yampolskiy@louisville.edu

Abstract

It is desirable to address accessibility issues within virtual worlds. Moreover, curbing criminal activities within virtual worlds is a major concern to the law enforcement agencies. Forensic investigators and accessibility researchers are gaining considerable interests in detecting and tracking avatars as well as describing their appearance within virtual worlds. Leinhart and Maydt have introduced a novel-set of Haar like features by extending the Viola Jones approach towards rapid object detection. We test this Haar cascade on human and avatar faces. Accuracy rates of 79% on human and 74% on avatar faces are obtained. The goal is to detect avatar faces in upright frontal face datasets and establish a baseline for future work in computer generated face recognition.

Introduction

Virtual worlds are gaining widespread momentum. They have a potential to transform the way the society operates. They bring in a sense of a "personal" digital space (Trewin et al. 2008) for the user by providing real time interaction with other fellow users and mirroring the real world activities. Communities, social groups, enterprises and institutions are all present in these virtual worlds. These virtual environments are gaining popularity across the globe and moving towards being an integral part of not only the Internet but also the society. An "avatar" is the user's virtual identity in these virtual worlds. The user can model its appearance to reflect either one's own personality or pose as someone else. Avatars can navigate within this virtual space by moving around buildings, flying in the air, swimming in the water or teleporting to different locations.

Destructive behaviors like terrorist activities and cybercrimes are reportedly on the rise within virtual worlds. There are reports of Al-Qaeda terrorists communicating and recruiting within the popular virtual world Second Life (SL) (Cole 2008), Second Life Liberation Army (SLLA) setting off virtual bombs to destroy virtual stores (Mandal & Ee-Peng 2008), American Apparel and Reebok's virtual store in SL being hit by virtual atomic bombs (Mandal & Ee-Peng 2008), etc. Identity thefts (Weinstein & Myers 2009) are a concern too. The lack of surveillance, presence of ungoverned spaces and the absence of strict rules monitoring the virtual society and its institutions have led to the growth of extremism and cyber-terrorism. Terrorist attacks, believed to be rehearsed in virtual worlds, are lethal as they can train using weapons identical to the real ones as well as build real life replicas of buildings and infrastructure.

Making these worlds accessible to a broad set of users is being gradually addressed. They are not easy for anyone to use the first time. Controlling an avatar requires simultaneous visual, audio, cognitive and motor abilities (Trewin et al. 2008). This results in a number of challenges to users with disabilities. Disabled gamers are showing a strong desire to be a part of these virtual worlds and envision themselves in their own avatars. Visually impaired users wish to detect and identify fellow avatars as well as other objects in their surroundings.

Criminal activities in these worlds are becoming a major problem for law enforcement agencies. Forensic experts are expressing interest in accurately and automatically tracking users and their avatars in these virtual communities. Detecting and recognizing avatar faces (Boukhris et al. 2011) will serve as one of the major component in building an artificial face authentication system (Gavrilova & Yampolskiy 2010) to help law officers track avatars within virtual worlds. The detected faces, saved to a database, will help to profile the avatars.

Profiling avatars based on their faces is a challenging and novel problem, contributing towards a new research direction in face recognition. Detecting avatar faces will address accessibility issues within virtual worlds, especially for visually impaired users, by describing the facial appearances of avatars in the vicinity by face detection.

Authenticating biological entities (human beings) is an essential and well-developed science, utilized to determine one's identity in today's modern society. However, avatar authentication (non-biological entities) is an issue that needs to be highlighted and addressed (Ajina, Yampolskiy & Amara 2010). A high degree of convergence between the real and virtual worlds has led to narrowing the distinction between the users and their avatars and applying security systems in these virtual spaces. To address the need for an affordable, automatic, fast, secure, reliable and accurate means of identity authentication Yampolskiy & Gavrilova define the concept of *Artimetrics* – a field of study that will allow identifying, classifying and authenticating robots, software and virtual reality agents (Yampolskiy & Gavrilova 2010).

Background

Avatar and human faces are similar as well as different. Both have consistent structure and facial component (eyes, ears, nose, etc.) locations. These similarities motivate the design of an avatar face detection framework based on principles similar to the human face detection system. Avatars have a wider range of colors than humans do that helps distinguish the two entities (Yampolskiy, Klare & Jain 2012). The goal is to detect avatar faces in the field of view of the user's avatar within virtual worlds that, along with face recognition, will help build a complete biometric authentication system for avatars. Currently available biometric systems are not equipped to deal with the visual and behavioral nature of artificial entities like avatars and perform poorly under such circumstances. Concerns over security and avatar identification are constantly voiced in virtual worlds (Yampolskiy & Gavrilova 2010).

Several challenges are involved in detecting avatar faces. They involve illumination, camera location, different skin color tones, pose, head rotation, etc. Certain preprocessing techniques such as geometric and color normalization may have to be applied (Yampolskiy, Klare & Jain 2012). In the context of investigating criminal activities in virtual worlds, we aim to examine some possibilities to authenticate avatars. These involve matching a human face to an avatar face when users upload their picture to model their avatars, matching the face of one avatar to another in a single as well as across multiple virtual worlds and matching a forensic sketch of an avatar to the avatar face (Yampolskiy, Klare & Jain 2012).

To the best of our knowledge, no prior research has been reported in the area of avatar face detection. However, there has been significant research in the domain of avatar recognition. The current state of art in virtual reality security, focusing specifically on emerging techniques for avatar authentication has been examined (Yampolskiy & Gavrilova 2010). Significant work has been conducted in recognizing avatar faces (Yampolskiy, Klare & Jain 2012). Research work has been carried out in the area of avatar facial biometric authentication (Ajina, Yampolskiy & Amara 2010). Daubechies wavelet transform and Support Vector Machines (SVM) are used to achieve artificial face recognition (Boukhris et al. 2011). In addition to these, there has been relevant research on robot emotion recognition (Yampolskiy & Gavrilova 2010). Intelligence Advanced Research Projects Activity (IARPA) is aiming to develop systems to observe avatar behavior and communication within virtual worlds to obtain insights into how real-life users in hostile cultures act and think (Yampolskiy & Gavrilova 2010). Another novel research direction is Avatar DNA, a patent pending technology by Raytheon. It focuses on providing authentication and confidentiality within virtual worlds by mixing real world biometrics of users with their avatar profiles (Yampolskiy & Gavrilova 2010).

In this paper, we focus on applying the face detecting OpenCV Haar cascade as an appearance based face detection technique. The method is based on an extended novel set of rotated Haar-like features, efficiently calculated by enriching the basic set of simple Haar-like features. Our test set comprises of human and avatar face datasets with varying backgrounds, illumination, rotations and face occlusions. The goal here is to obtain accuracy estimations by simply applying the cascade on each of these varying datasets.

The paper is organized as follows. We begin with an introduction to AdaBoost (Adaptive Boosting) learning algorithm, extended Haar-like features (Lienhart & Maydt 2002) and the OpenCV Haar cascade generation (Rhondasw 2009; Seo 2008). Next, we present the approach towards estimating the accuracies of applying the cascade on human faces and virtual world avatar datasets. The experimental results are described later. Finally, conclusions and directions for further enhancing the system performance are highlighted.

Algorithms

AdaBoost (Adaptive Boosting) Algorithm

AdaBoost algorithm, invented by Schapire and Freund (Schapire 1999), helped in solving many practical difficulties faced by the earlier boosting algorithms. It initially distributes a set of equal weights over a training set. After each round, the weak learning algorithm increases the weights for the incorrectly classified examples. This helps in focusing on the hard examples in the training dataset. "Ada" is a short form for adaptive as it adapts to the error rates of the individual weak hypothesis of each stage of the boosting process. Here the basic classifier is used extensively with the concept of stacking or boosting to constitute a very strong classifier. Several of these strong classifiers are subsequently connected into a cascade classifier to achieve the detection. The cascading levels determine the response of the system and the error rate. AdaBoost face detection techniques are based on the expansion of the Haar-like features, image integration and the cascaded classifier. For each image feature, a corresponding simple classifier is generated and the error relative to the current initialization error weight is evaluated. The classifier with the smallest error is chosen and added to the stage classifier. The weights of the samples are appropriately updated. If the sample is correctly classified, then the error is 0 or else it is 1. Finally, a stage classifier is obtained by combining the individual simple classifiers into a cascade. The algorithm is fast, easy and simple to implement, has no tuning parameters and requires no prior knowledge about the weak learner. Thus, it can be combined flexibly with any method to evaluate the weak hypothesis.

Extended Haar-like Features

These are a novel set of rotated Haar-like features, which yields a 10% lower false alarm rate as a face detector (Lienhart & Maydt 2002). It is an extension of the Viola Jones (Viola & Jones 2001) rapid object detection framework. It includes an efficient set of 45 degree rotated features that contribute additional domain knowledge to the learning process (Viola & Jones 2001). They can be computed rapidly at different scales.

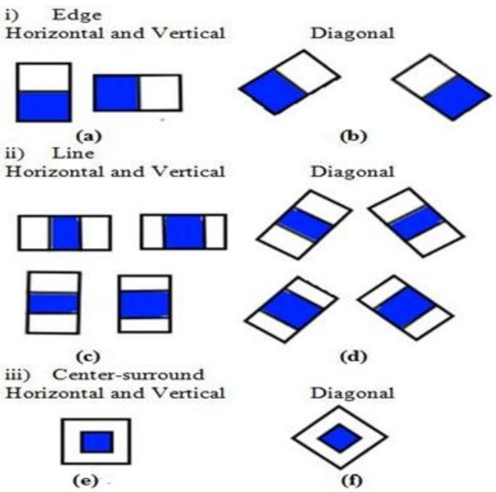

Figure 1: Simple Haar-like features. (a) and (b) are used to detect horizontal, vertical and diagonal edges respectively. Similarly (c) and (d) are used for lines and (e) and (f) for center-surround features. Shaded: Positive weights and Unshaded: Negative weights (Lienhart & Maydt 2002).

From Figure 1 we observe the 14 prototypes, which include four edge features, eight line features and two center-surround features. They are scaled independently in vertical and horizontal direction to generate a rich, complete set of features. The number of features obtained from each prototype is large and differs for each prototype (Lienhart & Maydt 2002).

Generating the OpenCV Haar Cascade

One of the available features of Intel's OpenCV (Intel) is face detection from images. Furthermore, it provides programs that are used to train classifiers for face detection systems, called Haar Training, to create custom object classifiers (Rhondasw 2009; Seo 2008).

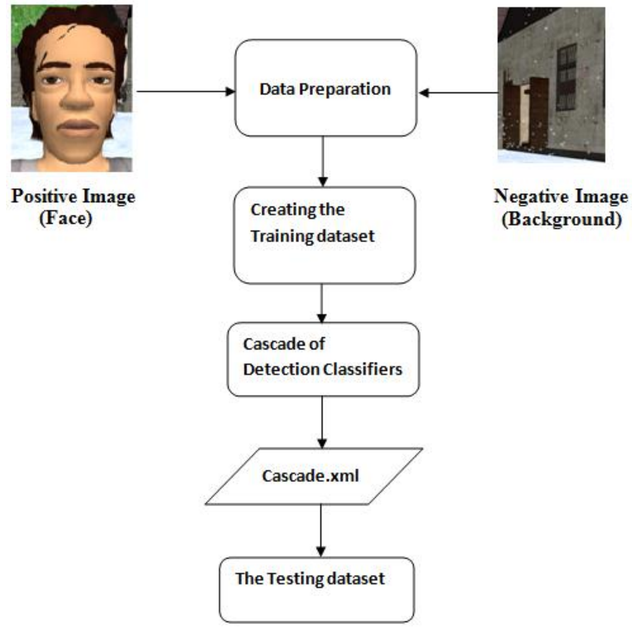

Figure 2: Flowchart for Haar training.

From Figure 2 we observe that the process of Haar training consists of the following steps:

<u>Data preparation:</u>

The process begins by gathering positive and negative datasets. The positive dataset contains images with the object of interest, i.e. the faces to be detected. The negative images are the ones that do not contain the object of interest, e.g. background images, non-face images etc. For real cascades there should be about 1000 positive and 2000 negative images. A general and acceptable positive-negative proportion is 1:2, but it is not a hard rule (Rhondasw 2009).

<u>Creating the training dataset:</u>

The prepared data now needs to be segregated into training and testing datasets. Here the training dataset is fed to a cascade of detection-boosted classifiers that yields the cascade xml file.

<u>Cascade of Detection Classifiers:</u>

Basic classifiers are put together to form stage classifiers which in turn are grouped together to form a cascade of stage classifiers. The series of such classifiers are applied to every sub-window of an image. A positive result from the first classifier stimulates the evaluation of the second classifier, which also has been adjusted to achieve high detection rates. A positive outcome from the second triggers the third and so on. Negative outcomes at any stage lead to rejection. Stages in the cascade are trained using AdaBoost and their thresholds are appropriately varied to minimize the false negatives. Thus, higher number of stages yields higher accuracies but leads to a decrease in performance time (Viola & Jones 2001).

<u>Cascade.xml:</u>

It is comprised of the various stages built due to the Haar training with the appropriate thresholds for each stage.

The Testing dataset:

The generated cascade xml file is evaluated on the testing dataset to get accuracy estimations of the results. The Haar cascade performs poorly on rotationally varying faces as well as occluded faces as it is based on object detection and it needs all the facial features (a pair of eyes, nose and mouth) to perform an accurate face detection.

Experiment

For the purpose of analysis of face detection techniques, we needed datasets with

- Faces with different head rotations
- Complex backgrounds
- Varying illumination
- Partially occluded faces

In our experiments, we used the following datasets:

Set 1 - Caltech

450 samples of human frontal face images from the California Institute of Technology were used(Weber). All images have the dimensions of 896 x 592 pixels. The dataset contains images from 28 subjects and 3 sketches with complex backgrounds and varying illuminations.

Set 2 – FERET

400 samples of human face images from the FERET ("The Color FERET Database") dataset were used. All images have the dimensions of 256 x 384 pixels. This dataset contains images for 52 subjects. Each subject is represented by 7-8 images with random head rotations varying from 67.5 degree to 15 degree rotations in both left and right directions, plain background and slightly varying illuminations.

Set 3 – Avatar

Avatar faces from the popular online virtual worlds, Entropia Universe and Second Life, were used in the avatar dataset. A scripting technique was designed and implemented to automatically collect the avatar faces using AutoIT as well as Second Life's Linden Scripting Language (LSL) (Oursler, Price & Yampolskiy 2009; R.V. Yampolskiy & Gavrilova 2010).

The dataset is subdivided into 3 parts with the number of samples indicated in the parenthesis next to it: Female avatars from Entropia Universe with complex backgrounds (150), Male avatars from Second Life with a regular or plain background (150) and Male avatars from Second Life with complex backgrounds (150). The dataset contains the avatar faces with a set of five images per avatar from different angles with complex backgrounds and varying illuminations. Figure 3 shows a sample of this dataset with different random head rotations and a complex background.

Figure 3: Examples of different face images of the same subject from the Second Life avatar dataset with a complex background. Each image corresponds to the different head rotations while facing the camera. The frontal image is (a). (b) to (e) represent the same avatar with varying head rotations.

The experiment involved using a single OpenCV Haar cascade, *haarcascade_frontalface_alt.xml*, on these datasets and evaluating the algorithm's performance on each of them. There is no training involved here. The experiment relies solely on the accuracy of this cascade on the human and avatar datasets. Basically, a classifier trained on human faces is used to detect avatar faces without any special preprocessing.

Initially, a code similar to the one found in (Hewitt 2007) was written in Microsoft Visual Studio 2008 using the OpenCV libraries (Intel). This code helps in detecting faces within images using the OpenCV Haar cascade. The cascade was applied on the dataset images. Promising results were obtained and a comparitive analysis was drawn.

Results

A summary of the results, obtained by executing the above OpenCV Haar cascade on all three datasets, are shown in Table 1. These results include images that belong to multiple categories. A brief description of each category is given below.

Background: The nature of the background, which can be either plain or complex.

Positive: Signifies a positive detection with a face accurately detected.

Background Faces (BF): The background faces (if any) other than the main subject that are detected.

False Positives (FP): Non–face objects that are incorrectly detected as faces.

False Negatives (FN): Background faces that are not found by the detector.

Zero Detections (ZD): The face of the main subject as well as those in the background (if any) that are completely undetected by the detector.

Accuracy: The positive detections.

Error Rates: Comprises of false positives, false negatives and zero detections.

Average Accuracy and Error Rates: The average of Accuracy and Error Rates obtained.

A: Images with BF’s detected without FN
B: Images with BF’s detected with FN
C: Images with both FP’s and FN’s
D: Images with only FP’s
E: Images with both ZD’s and FP’s
F: Images with only FN’s
G: Images with only ZD’s

Table 1: Face detection results for all three datasets.

Dataset	Human		Avatars		
Type	***Caltech***	***FERET***	***Entropia***	***SL (Male)***	***SL (Male)***
Background	Complex	Plain	Complex	Plain	Complex
Positive	444/450	235/400	114/150	79/150	140/150
Background Faces (BF)	11/25	0/0	0/0	0/0	0/0
	A **B** 7/11 4/11				
False Positives (FP)	11/450	0/400	0/150	0/150	0/150
	C **D** **E** 2/11 8/11 1/11				
False Negatives (FN)	12/450	0/400	0/150	0/150	0/150
	B **C** **F** 4/12 2/12 6/12				
Zero Detections (ZD)	6/450	165/400	36/150	71/150	10/150
	E **G** 1/6 5/6				
Accuracy (%)	444/450 = 98.6	235/400 = 59	114/150 = 76	79/150 = 52.66	140/150 = 93.33
Accuracy (Average %)	(98.6 % + 59 %) / 2 = 78.8		(76 % + 52.66 % + 93.33 %) / 3 = 74		
Error Rates (%)	29/450 = 6.4	165/400 = 41	36/150 = 24	71/150 = 47.33	10/150 = 6.66
Error Rates (Average %)	(6.4 % + 41 %) / 2 = 23.7		(24 % + 47.33 % + 6.66 %) /3 = 26		

Set 1 - Caltech

On the 450 human face images, the OpenCV Haar cascade yielded an accuracy rate of 98.6% with an error rate of 6.4%.

Set 2 - FERET

On the 400 human face images, the OpenCV Haar cascade yielded an accuracy rate of 59% with an error rate of 41%. The error rate being high is mainly due to the poor performance of the cascade on rotationally varying faces as well as face occlusions.

Set 3 - Avatar

Facial images from the three separate avatar datasets are fed to the OpenCV Haar cascade individually. Accuracies obtained are 76% (Female avatar-Entropia, complex background), 52% (Male avatar-Second Life, regular background), 93% (Male avatar-Second Life, complex background) and error rates of 24% (Female avatar-Entropia, complex background), 48% (Male avatar-Second Life, regular background) as high as 7% (Male avatar-Second Life, complex background). From the results we observe that the cascade performs very well on the Male avatar-Second Life (complex background) but not so good on the Male avatar- Second Life (regular background).

Figure 4 shows a set of bar graphs for the results obtained on the FERET and the avatar datasets.

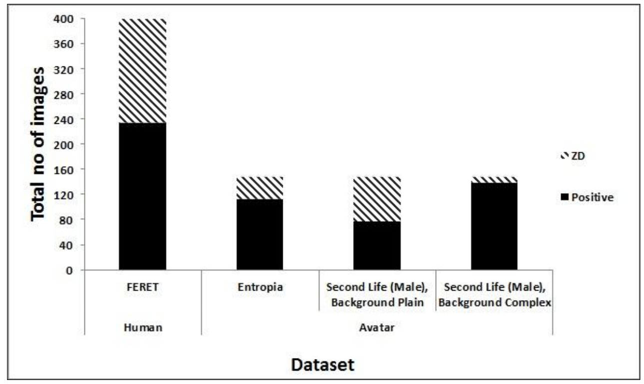

Figure 4: Results from the FERET and the avatar datasets.

A set of sample detections are shown in Figure 5.

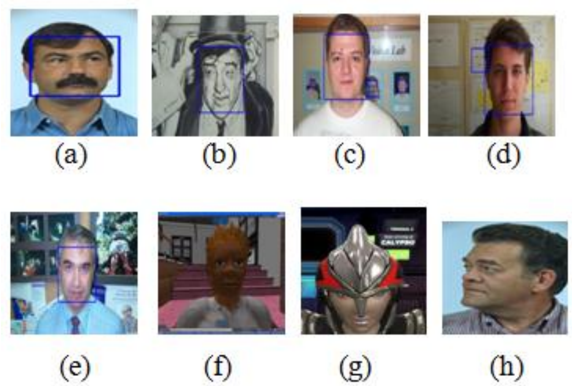

Figure 5: (a) FERET dataset – Positive detection (b) Caltech dataset – Positive sketch detection (c) Caltech dataset – Three background faces detected (d) Caltech dataset – False positive (e) Caltech dataset – False negative, background face at lower right corner (f) Male avatar: Second Life, complex background dataset – Zero detection due to poor illumination (g) Female avatar: Entropia, complex background dataset – Face occlusion (h) FERET dataset: Zero face detection due to 90 degree face rotation to the left.

Conclusion

This paper evaluates a standard OpenCV face detection algorithm, which utilizes a Haar cascade on a set of human facial images as well as virtual world entities: avatars, which are rapidly becoming part of the virtual society. Accuracy estimations reported from each dataset are compared. Good results are obtained, but poor performances are recorded for facial images involving head rotations, poor illumination and face occlusions. The average accuracy detection rates on human faces is 79% whereas for avatar faces it is 74%, which is quite good given the quality of the picture, appearance of the avatar as well as background and surrounding regions. The varying accuracy rates are due to the difference in the count of head rotations, face occlusions, illumination conditions and plain/complex backgrounds between the datasets.

Potential directions for future research involve improving the existing algorithm to achieve better accuracies by fusing the current algorithm with Local Binary Patterns (LBP). We will extend this approach over a larger dataset to yield more accurate evaluations and better results. Further, unifying this face detection technique with face recognition to build a complete face authentication system capable of authenticating biological (human beings) and non-biological (avatars) entities is the direction in which we have set our sights on.

References

Ajina, S., Yampolskiy, R. V. & Amara, N. B. (2010, Jul 1-2). *Authetification de Visages D'avatar [Avatar Facial Biometric Authentication]*. Paper presented at the Confere 2010 Symposium. Sousse, Tunisia.

Boukhris, M., Mohamed, A. A., D'Souza, D., Beck, M., Ben Amara, N. E. & Yampolskiy, R. V. (2011, July 27-30). *Artificial human face recognition via Daubechies wavelet transform and SVM.* Paper presented at the 16th International Conference on Computer Games (CGAMES). Louisville, KY.

Cole, J. (2008). Osama Bin Laden's "Second Life" *Salon*, from http://www.salon.com/news/opinion/feature/2008/02/25/avatars

The Color FERET Database. Available from NIST The Color FERET Database http://face.nist.gov/colorferet/

Gavrilova, M. L. & Yampolskiy, R. V. (2010, Dec). State-of-the-Art in Robot Authentication [From the Guest Editors]. *Robotics & Automation Magazine, IEEE, 17*(4), 23-24.

Hewitt, R. (2007). Seeing With OpenCV, Part 2: Finding Faces in Images. *Cognotics: Resources for Cognitive Robotics*, from http://www.cognotics.com/opencv/servo_2007_series/part_2/index.html

Intel. Open Computer Vison Library, from http://opencv.willowgarage.com/wiki/ http://sourceforge.net/projects/opencvlibrary/files/opencv-win/2.1/

Lienhart, R. & Maydt, J. (2002, Sept 22-25). *An extended set of Haar-like features for rapid object detection.* Paper presented at the International Conference on Image Processing. Rochester, NY.

Mandal, S. & Ee-Peng, L. (2008, May 12-13). *Second Life: Limits Of Creativity Or Cyber Threat?* Paper presented at the 2008 IEEE Conference on Technologies for Homeland Security. Waltham, MA.

Oursler, J., Price, M. & Yampolskiy, R. V. (2009, Jul 29-31). *Parameterized Generation of Avatar Face Dataset.* Paper presented at the CGAMES '2009: 14th International Conference on Computer Games: AI, Animation, Mobile, Interactive, Multimedia, Educational and Serious Games. Louisville, KY.

Rhondasw. (2009). FAQ: OpenCV Haar Training, from http://www.computer-vision-software.com/blog/2009/06/opencv-haartraining-detect-objects-using-haar-like-features/ http://www.computer-vision-software.com/blog/2009/11/faq-opencv-haartraining/

Schapire, R. E. & Shannon Laboratory ATT. (1999, Jul 31- Aug 6). *A brief introduction to boosting.* Paper presented at the Sixteenth International Joint Conference on Artificial Intelligence. Stockholm, Sweden.

Seo, N. (2008). Tutorial: OpenCV Haartraining (Rapid Object Detection with a Cascade of Boosted Classifiers Based on Haar-like features) from http://note.sonots.com/SciSoftware/haartraining.html

Trewin, S., Laff, M., Cavender, A. & Hanson, V. (2008, Apr 5-10). *Accessibility in virtual worlds*. Paper presented at the CHI '08 Human factors in computing systems. Florence, Italy.

Viola, P. & Jones, M. (2001, Dec 8-14). *Rapid Object Detection using a Boosted Cascade of Simple Features*. Paper presented at the Computer Vison and Pattern Recognition. Kauai, HI.

Weber, M. A frontal face dataset collected by Markus Weber at the California Institute of Technology. Available from California Institute of Technology Bg-Caltech http://www.vision.caltech.edu/Image_Datasets/faces/README

Weinstein, J. & Myers, J. (2009, Dec 2). Avatar Identity Theft Prompts Review of Gaia and Virtual World Rules, from http://www.jackmyers.com/commentary/guest-mediabizbloggers/11193066.html

Yampolskiy, R. V. & Gavrilova, M. (2010, Oct 20-22). *Applying Biometric Principles to Avatar Recognition.* Paper presented at the International Conference on Cyberworlds (CW2010). Singapore.

Yampolskiy, R. V., Klare, B. & Jain, A. K. (2012, Apr 23-27). *Face Recognition in the Virtual World: Recognizing Avatar Faces*. Biometric Technology for Human Identification IX. SPIE Defense and Security Symposium. Baltimore, MD. (In Press).

D'Souza, D., & Yampolskiy, R. V. (2010). Avatar Face Detection using an Extended Set of Haar-like Features. [Journal Abstract]. *Kentucky Academy of Science, 71*(1-2), 108.