# SUBTLE PRESENCE: DESIGN AND IMPLEMENTATION OF USER CENTRIC CONTENT DELIVERY USING BIOMETRIC DATA CAPTURE AND INTELLIGENT ANALYSIS


## WILLIAM PENSYL



Explores the user centric delivery Media content, using biometric data capture, intelligent analysis of facial data, age, gender and other forms of data that can be directly captured in a non-invasive manner. The system has an inherent intelligence that is ambient, and ubiquitous – allowing for interpretation of a wide variety of stimuli and that can be easily collected.


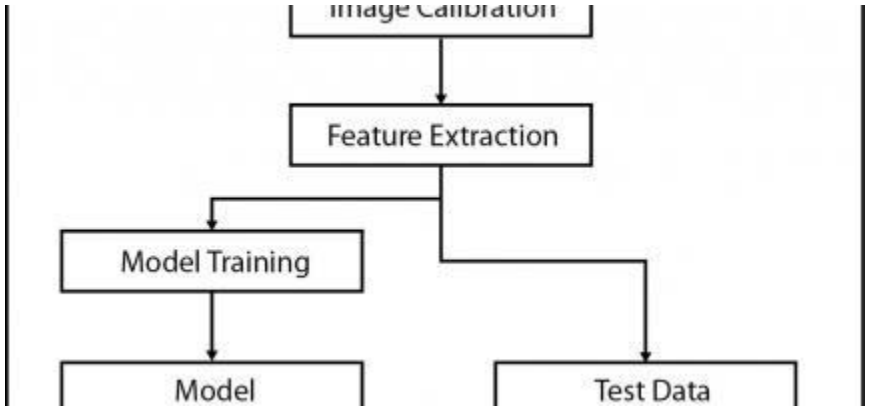


*Figure 1. Processing Flow Diagram of HiPOP System*

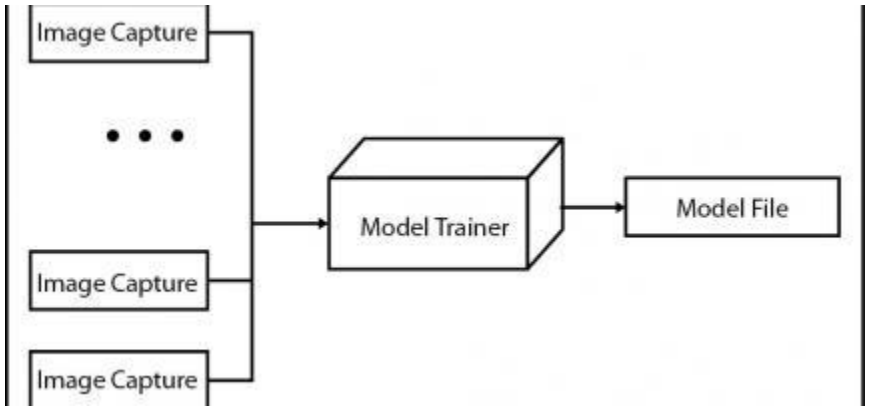


*Figure 2. Data Training from Datasets*

*Figure 3. Subtle Presence, presented in the 2011 Sarajevo International Winter Festival.*

This paper presents design and implementation of user centric content delivery using biometric data capture and intelligent analysis. We describe feasibility and successful implementation of responsive information delivery tools prefiguring facial and biometric data to cue advertising, social communication or culturally relevant user experiences. Initially designed for marketing content in public spaces, the content delivery can vary depending location and population demographics. Various forms of data, captured in a non-invasive manner, including facial images, height, body type, age, gender and aspects of mood can be used determine appropriate media delivered based on these personal attributes. To alter the type of media content presented in advertising, in retail environments, exhibition installations or public spaces, it may prove useful for media designers or systems designers to be able to assess certain personal attributes of the viewer in the space. We refer to this system as HiPOP – A High Impact Point of Purchase content delivery system.

The system uses ambient and ubiquitous intelligence to detect a face, calibrate the image and extract features to classify and determine personal attributes. After testing for gender and age group, media content is presented based on these attributes. Such a system must have an inherent intelligence and modest decision making ability that is ambient and ubiquitous – allowing for interpretation of a variety

of stimuli. The intelligence must allow meaningful responses to visual and sensor cues. There are many possible applications for the implementation this system, including, information delivery for targeted advertising, social communication in public environments. It also can be used to create socially engaging integrated media artworks within architectural and exhibition spaces. This allows viewers to engage in aesthetic experiences that are subtly responsive to their personal physical attributes and moods.

Project overview

There are a number of developments that are required for this type of information delivery. These build upon previously published biometric data capture techniques. The significance of the work lies in the development of an autonomous, intelligent system that can deliver user specific information based on the collected data of a viewer's gender, height, weight and other cues that allow for a definition of a user profile. One critical aspect to such a system is an ambient and non-invasive data capture with a naturalistic, subtle response to the user.

There has been much research on capturing user biometric information, gaze detection, posture, and these have been used to create interactive systems that allow for more natural interactions in games, interactive and environments. However, no application has been developed for point of purchase environments. In this project, we want to focus on getting useful information delivered for viewers in this specific application. Furthermore, we utilize hardware/sensor and vision systems to increase the range and accuracy of information delivery, providing better solutions for advertising, customer support, and open these platforms for the creation of interactive artistic installations.

**System overview**

The system processes the images captured through three modules: a "Detection Model," a "Data Training Model," and a "Demo Showing Model." The Detection Model algorithm detects a face, calibrates the image and extracts features using OpenCV, Haar-like application and LibSVM to classify and determine gender. [11], [8] An AdaBoost learning algorithm boosts classification performance. [6], [7] [8] The Data Training Model uses a cascading classification method and a LibSVM to train analysis of data and generate a final data model file. [8] The Demo Showing Model manages windows for the system and audience delivery content. The detection result is shown in face detection window and scene view window. The content images or steaming video is shown in a second display monitor.

The Detection model uses sensing and vision technology that captures a video stream, an algorithm that analyzes and identifies if a face is present, then compares the detected face to a library of defined face images organized by gender and age. This allows for the determination of characteristics including gender and age within a set of age groups. The method of detection and classification uses OpenCV Haar-like Features to find a face rectangular within the space where video camera is pointed. [1] A LibSVM is used to classify and determine the final result in gender and age bands. The results of the content delivery selected are targeted to groups that are more easily defined. [8] Gender is easier to detect than age. The fine distinction between age within the small child group, or with the adult groups is more difficult to accurately determine.

The HiPOP System is detailed in the processing flow diagram in Figure 1. The image is captured by any consumer grade webcam and passed to the face detection and image calibration module. The features are extracted via a very fast feature evaluation using Haar-like functions and AdaBoost [5] to increase to focus in a small set of critical features.

The Data Training Model is necessary to ensure the software algorithms can compare the data set of images with the library and the images captured can be calibrated to increase accuracy in the final result.

Using the LibSVM data file the system is trained to analyze the data, to classify the faces according to a set of grouped ages and genders, and to generate a final data model file. Different data models can be created using different data sets. The Demo Showing Model shows the data detection results and sends the data to playback system, either a QuickTime or Windows Media Player.

This implementation relies on published work on Haar-like Features [10]. Haar-like features are digital image features used in object recognition. They owe their name to their similarity to Haar wavelets. Viola and Jones implemented very high frame rate object detection with only the information in a single grey scale image, using rectangular Haar-like features. A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called two-rectangle feature. Viola and Jones [1] also defined three-rectangle features and four-rectangle features in the object detection framework. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a two-rectangle feature can indicate where the border lies between a dark region and a light region.

The Viola-Jones object detection framework has three steps to extract features and classification. Rectangular Features are enclosed within a detection rectangle. The area rectangle is divided horizontally and vertically. The value of the divided rectangles is determined and the differences in features are found. Viola and Jones refer to the method employed as the “Integral Image.” This first step is used to determine the rectangular features in an intermediate representation. The Integral image allows the number of iterations in processing the image to be limited, thus increasing the speed of the feature extraction. The second step uses a variant of the AdaBoost [5] to select a small set of features and train the classifier. First, collect a group of pictures, some with human faces and some without. For each image, give the image a weight (1/m [human face], 1/l [non-human face]). And then extract “T” features from a lot of images. The processes for extraction, repeated T times are: first, standardize the weight, the sum is 1; second, pick up the feature that has smallest error; third, record the parameter where the error is smallest; and finally, refresh the image weights. After the above processes, we can decide if this image contains a face. The third step uses a set of Cascade Classifier Functions to compare iteratively against at library of images that are previously classified as faces, and within a gender or age group. Through testing and comparison of the captured face, the result is determined, within a margin of error. A larger library for comparison increases the accuracy of the result. There is a trade off in the speed of processing due the comparison of larger ldatasets. For purposes of this implementation, the determination needs to be very fast. To maintain a fast response, we work with a limited set of images in the library. The first step determines the results in male or female gender. Following, the face images are classified via the cascade classifier function to determine age grouping. In our implementation, the age groups are limited to child, teen, young adult, adult and seniors. The system implementation also includes the detection of a smile, adding the potential for determination of a certain amount of mood within the face of the individual viewed.

To increase the accuracy of the result, determining the age and gender, the data training models Support Vector Machines (SVMs) supervise the learning methods used for classification and regression. SVMs are classifiers the extract the results belonging to one of two categories. The SVM training algorithm builds a model that predicts whether a new example will fall in to one or the other category. The use of the SVM increases the accuracy of the resulting detection.

**Implementation**

The HiPOP system was designed for environments where narrow cast media is delivered in an environment; such as a retail outlet point of purchase display is maintained. Other locations where a system such as this can be employed are in any location where marketing or advertising content is presented in a public space to one person or to small groups of people. It is common that we see narrow cast information displays in point of purchase locations, elevators, entry foyers, and even in mass transit trains and subway cars. For such content delivery to have a maximum impact the content can be targeted to the person that is looking at the monitor. In these circumstances, the system can play content that is of interest to the person there. This targeting of content may increase apprehension and response by the viewer. In most cases of content delivery in these types of displays, the content is simply streamed without consideration for who may be viewing it. The content may be completely off target. For example, in a point of purchase display, the customer could be an older female. Yet the content of the media may for a product that is commonly used only by young males. Extreme examples of this kind of disconnect occurs when the product marketed is related to a life style or behavior that may be considered inappropriate by the viewer. This could lead to a negative experience, and decreasing participation by the consumer. If the content can be targeted to the individual, then interest and participation can be maximized.

One unique feature of the HiPOP system is that the viewer is never aware that the media content is focused to them. The system simply displays the content in a seamless manner, without any indication that a detection of personal attributes is made. The system uses a web cam mounted near a display screen. The camera captures the faces in the space and detects personal attributes. Using the techniques described above, the system determines and classifies the person and then displays the targeted content. There is a limit to the accuracy of the detection. In ideal circumstances, gender detection is 80 to 85% accurate. However, since the system is designed to display the results without the knowledge of the viewer, the viewer does not experience any negative responses. They are unaware that the content displayed is the result of an inaccurate detection of personal attributes.

Another use of the system is in subtle, conceptual and artistic interactive media installations in public spaces. This is useful for interactive or streaming media content in exhibition to be displayed and altered based on the viewer's personal attributes. In Figure 3, the installation depicted alters a still life painting and changes it over time. The result of face detection alters the image according to the viewer age and gender.

In upcoming installations, importance is placed on the detection of mood. As the system software can easily detect facial expressions, the altering the content to match the mood of the viewer can easily be attempted. Current work explores how images can be shifted due to the mood displayed in the viewer in exhibition spaces. Detecting reaction to content and altering the content stream in ways that increase satisfaction is possible. Alternatively, we will explore how mood of the viewer may be affected by the images. There may be ways that various images types can be used to evoke mood shifts within the viewer. In the current installation we are detecting the smile in the viewer's faces. We can track the length of time the viewer is smiling, or other limited facial expressions. The longer one smiles, or alters their body posture; the vibrancy and saturation levels of an image can be adjusted. Other type of images can be streamed based on the visual cues collected. This new work requires "training" of the model to classify the images detected according to the set of groups defined. The library of images is stored in a database with key identifiers associated with mood: smiling, eyebrow position, eye wideness, body poster, and slope of shoulders, head tilt, or other detectable and identifiable cues.

**Conclusion**

Such systems can provide valuable mechanisms for delivery of media content in public places. One goal in advertising and marketing is increasing interest and engagement of the viewer. By targeting the content more closely to the person based on their own specific attributes, engagement in the content can be maximized. For artistic or conceptual installations, specific goals can be achieved through interaction that is responsive, yet unobtrusive. In the system we have implemented, there is no data or personal information is actually stored by the system. The system simply detects a face or personal cues and defines a classification of the person based on age and gender, and plays a piece of media content. This system creates a subtle responsive interaction that is unobtrusive, yet provides valuable media content that has the potential to increase viewer's engagement in meaning ways.

***References and Notes:***


*1 Viola, P., Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*

*2 Burges C.J.C., (1998) A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2*

*3 Sun Z., Yuan X., Bebis G., Louis, S.J., (2002). Neural-Network-Based Gender Classification Using Genetic Search for Eigen-Feature Selection. IEEE International Joint Conference on Neural Networks.*

*4 Phillip Ian Wilson P.I., Fernandezj., (2006). Facial feature detection using Haar classifiers, Journal of Computing Sciences in Colleges, Vol. 21, Issue 4*

*5 Freund Y, Schapire R. E. (1995).A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt '95, pages 23–37. Springer-Verlag,*

*6 Schapire R.E., Freund Y., Bartlett P., Lee W.S., (1997) Boosting the margin: A new explanation for the effectiveness of voting methods. Proceedings of the Fourteenth International Conference on Machine Learning*

*7 LibSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm/*

*8 FERET Database: http://www.nist.gov/humanid/colorferet*

*9 The MPLab GENKI-4K Dataset: http://mplab.ucsd.edu,*

*10 Haar-like features: http://en.wikipedia.org/wiki/Haar-like_features*