Computer Vision and Digital Signage

Borut Batagelj
University of Ljubljana,
Faculty of Computer and
Information Science
Computer Vision Laboratory
Tržaška 25, 1000 Ljubljana,
Slovenia
borut.batagelj@fri.uni-lj.si

Robert Ravnik
University of Ljubljana,
Faculty of Computer and
Information Science
Computer Vision Laboratory
Tržaška 25, 1000 Ljubljana,
Slovenia
robert.ravnik@fri.uni-lj.si

Franc Solina
University of Ljubljana,
Faculty of Computer and
Information Science
Computer Vision Laboratory
Tržaška 25, 1000 Ljubljana,
Slovenia
franc.solina@fri.uni-lj.si

ABSTRACT

Digital signage in combination with computer vision offers an effective platform for out-of-home advertisement which can offer accurate audience measurement data. Such intelligent digital signage systems can also adapt the displayed contents to the actual audience in real time and even enable simple non-contact human computer interaction.

In this paper we would like to show how digital signage can be made much more effective by using computer vision technology. Computer vision methods for face detection, classification and recognition of people's behavior in front of digital displays can provide some objective data about the people (demographic data such as age and sex for example) who have observed the messages displayed. Furthermore, computer vision can be used for a non-contact way of user interaction with the displayed content.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Experimentation, Human factors

Keywords

Digital Signage, Computer Vision, Marketing, Advertisement, Face detection, Face classification

1. INTRODUCTION

Digital signage is the newest player in the world of out-of-home advertising. The term digital signage refers to screens both large and small that are used to show content and advertising. The screens are usually networked to a main content server which can usually be administered from anywhere in the world where an internet connection is available.

The benefits of digital signage are clear and are being realized by thousands of retailers, government institutions, and educators worldwide. The dynamic digital signage is eye catching, and can be used to up-sell at point-of-sale where add-on purchases might not normally occur. Digital signage also serves to modernize the retail area [4].

The digital signage phenomenon has become more popular as the cost of equipment has fallen. Large plasma or LCD screens are more affordable today than ever before. The servers and PCs that are used to administer the content are also extremely affordable. This attainability has lead to explosive growth in the digital signage sector over the past two years. Because of the affordability, companies both large and small have been taking advantage of this trend. Digital signage has also become common place in hospitals, schools, post-secondary institutions, government, airports, shopping malls, and financial institutions, to name a few.

Methods for face detection and recognition have recently gained a significant progress, especially during the past few years. Earlier face detection techniques could only handle single frontal faces in images with simple backgrounds, while state of the art algorithms can detect and track faces and their poses in cluttered backgrounds [9]. Beside face detection and recognition, facial expression analysis [6], age and ethnic [3] classification has been an active research topic. With new pattern recognition methods we can further analyze and classify detected faces.

In the next section, we introduce the basic idea of Intelligent Digital Signage and technical background how to build such an itelligent display. Section 3 describes our system for analysis and classification of faces and technical details of usage such a system for digital signage purpose. The final Section conclude this work with the list of usefull data information which we can get from our system and there practical use in the marketing area.

2. THE INTELLIGENT DIGITAL SIGNAGE

With today's technological advancement, businesses, advertisers, and marketers alike are all looking for new, attention grabbing, cost effective ways to advertise their products and catch the eye of potential customers. By deploying an inexpensive video sensor on the top of the screens and taking advantage of the extra, unused computing power of the associated signage players, the system can be made intelligent by measuring and adapting to the actual audience in real time (Figure 1). Computer vision methods can detect, categorize and analyze the people's faces and according to the results one can estimate the number of current viewers, their distance from the screen, their demographics and their

MIAUCE 2008 Chania, Crete, Greece

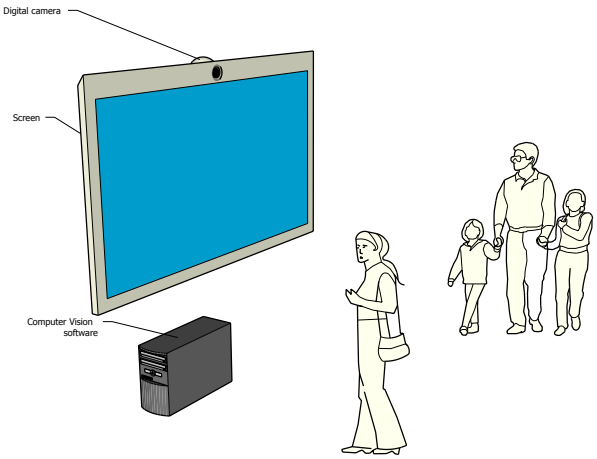

Figure 1: Example of an Intelligent Digital Signage setup

behavior.

An intelligence digital signage system could display therefore just the advertisement targeted at the audience which is actually present at the moment. Screen time could therefore be billed per qualified impression. Due to the inability to precisely target prospective clients, the equitable advertising billing system is a very real concern for all advertisers. An intelligent digital signage system can give the right message to the right audience in real time. With the help of computer vision methods the real-time audience measurement data, message targeting, interactivity and accountability can become a reality for today's digital signage. This new type of advertising is an attempt by companies to improve their Point of Purchase Advertising techniques so these types of advertisements are especially prevalent in stores where the product being advertised is being sold. With the camera hidden behind the product and the analysis of facial expression of potential customer one can detect his/her reaction on specific product. To make the advertisement signs even more attractive one can introduce also some simple interaction with the viewers. By defining "hot spots" in the space before the screen/camera the viewer can select specific multimedia contents just by positioning himself in the right spot. For interaction one can use also gesticulation or other behavior of viewers.

In the context of our interactive fine art installation entitled "15 seconds of fame" we displayed portraits of randomly selected viewers that were graphically converted in pop-art style [7, 8]. The general setup was actually quite similar to an intelligent signage display (Figure 2). We observed that the behavior of the audience was definitely influenced by the interactive nature of the installation. Even people without any prior information on how the installation works quickly realized that the installation displays portraits of people who are present at the moment. Suddenly, subtle staging takes place in front of the installation to get one's most favorable image on the screen, especially since the audience does not know the exact moment when the next picture is taken. But getting a share of that fame and seeing one's own portrait on the wall proved to be quite elusive if several people were in

Figure 2: Children having fun in front of the interactive art installation "15 second of fame"

the audience since the featured face was selected randomly out of all detected faces. People who would step right in front of the installation, somehow trying to force the system to select them for the next 15 second period, would be more often disappointed, seeing that somebody else way back or on the side was selected instead. A mini reality show in the manner of Big Brother would sometimes take place with open (Figure 2) or more subdued competition for media attention, illustrating the theatricalization and the need of self-presentation in all spheres of life [8].

3. THE FACE CLASSIFICATION SYSTEM

Research in the area of face and gesture analysis has been driven mostly by the demands in the area of security, biometrics and surveillance where the efficiency and accuracy must be as high as possible. The accuracy of the same techniques in the context of digital signage is not as critical. More important are some other issues such as privacy in particular. Since digital signage systems are employed also in public places the identities of individuals should not be revealed. Therefore the actual faces of viewers should not be recorded; only the qualitative data associated to the detections and classifications should be recorded and sent to the server. Additionally, in the images the viewer's faces can be blurred for privacy reasons.

We are developing an extended system for face recognition that can also analyze and classify faces [1, 2]. We are testing this system for use in a digital signage context. The basic face recognition system consists of four modules: detection, alignment, feature extraction, and matching, where localization and normalization (face detection and alignment)

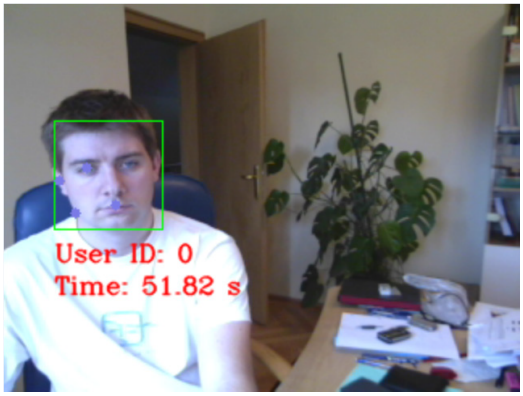

Figure 3: Face detector detects and tracks the faces in front of the screen and measures the time of their appearance.

are processing steps before face recognition (facial feature extraction and matching) is performed.

The face detection module segments the face areas from the background. The detected faces can be tracked using a face tracking component. Face alignment module is aimed at achieving more accurate location of faces since face detection provides coarse estimates of the location and scale of each detected face. Facial components, such as eyes, nose, and mouth and facial outline, are located; based on the location points, the input face image is normalized with respect to geometrical properties, such as size and pose, using geometrical transforms or morphing. The face is usually further normalized with respect to the photometrical properties such illumination and gray scale. After a face is normalized geometrically and photometrically, feature extraction is performed to provide effective information that is useful for distinguishing between faces of different persons and stable with respect to the geometrical and photometrical variations. For face matching, the extracted feature vector of the input face is matched against those of enrolled faces in the database; it outputs the identity of the face when a match is found with sufficient confidence. On the feature extraction level we extract features which exactly define facial attributes which are important for distinguishing between different classes of ages, gender, race group (Africans, Europeans, Asians) and even different expressions (disgust, fear, joy, surprise, sadness, anger). Detailed analysis can recognize facial hair (beard, mustache), makeup, eyeglasses, headgear, shawl, and hair (style, color, and length). With texture analysis we can detect if the skin on the face is wrinkly, smooth or even pimpled.

The system for digital signage consists of two main parts. The detector captures the frame and computes when and how long a face is present in front of the screen (Figure 3). Additionally, it also computes the distance from the face. Face classification system then classifies and analyses the detected face. The face detector is based on the AdaBoost [9] learning-based method because it is so far the most successful in terms of detection accuracy and speed. AdaBoost is used to solve the following three fundamental problems:

1. learning effective features from a large feature set;
2. constructing weak classifiers, each of which is based on one of the selected feature set; and
3. boosting the weak classifiers to construct a string classifier.

Weak classifiers are based on simple scalar Haar wavelet-like features, which are steerable filters. We use the integral image method for effective computation of a large number of such features under varying scale and location, which is important for real-time performance. Moreover, the simple-to-complex cascade of classifiers makes the computation even more efficient, which follows the principles of pattern rejection and coarse-to-fine search. When the face is detected we track it with the Lucas-Kanade method [5], which is a two-frame differential method for optical flow estimation. Considering time and space locality we achieve the satisfactory accuracy. When the detector loses the detected face it sends the captured data in XML format to the server. Transmitted data is transformed to proper format and saved in a database. The second part of the system is a web application, which is able to generate the reports from the database data (Figure 4). For the web server we use Apache and Linux. All dynamic reports are made with help of a PHP script in conjunction with Adobe Flex framework. The system is now ready to be moved from a laboratory setting to a real environment for testing.

4. CONCLUSIONS

An intelligent digital signage system can produce real-time data, which can be used in several ways:

- using computer vision methods analyze the audience in front of the display in each given time period estimating the number of viewers and their profiles (sex, age, mood or activity),
- use the audience analysis to determine typical data which is important for advertisers, such as:
  - determine the precise count of actual viewers,
  - compute various aggregate indices on viewership such as dwell time, attention time, “face minutes” etc.,
  - determine precise viewership demographics,
  - determine precise correlation between viewership and content,
  - the advertising can bill campaigns by the actual impression,
- use the audience analysis in real time to:
  - show specific messages according to predefined rules, reacting to viewer profiles (for example, select the best performing messages for a particular gender or age group) or to specific behaviors such as approaching or turning away from the screen, to the distance of the viewer from the display etc.,
  - activate the sound only when customers are watching the screen to reduce employee fatigue,

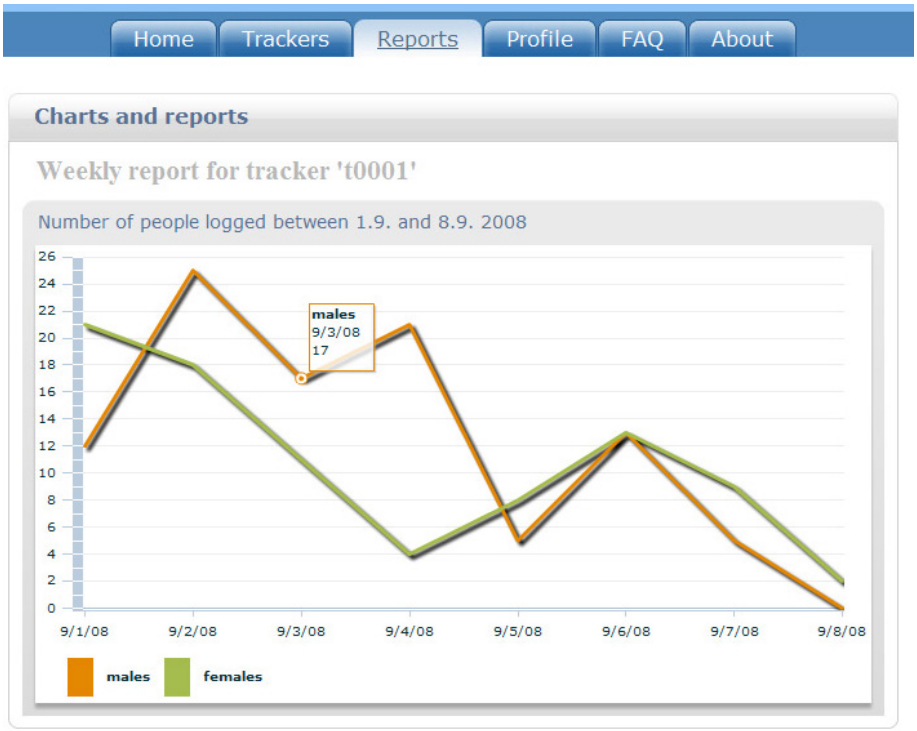

Figure 4: Example report on how many women and how many men watched the advertisement in the selected week.

– estimate the opportunity to see.

The intelligent signage system that adapts to the audience in real time should not over react. The audience and its behavior should probably be classified into a small number of classes for which particular rules could be applied. How such an intelligent signage system should be structured so that some general rules about advertising could be followed and that an administrator of the system could easily adapt its behavior to different contents is an interesting research problem.

5. REFERENCES

[1] B. Batagelj, F. Solina. Face recognition in different subspaces : a comparative study. In A. Fred, A. Lourenco (Eds.), Pattern recognition in information systems: proc. 6th Int. Workshop on Pattern Recognition in Information Systems, PRIS 2006 in conj. with ICEIS 2006, Paphos, Cyprus, Instic Press, pages 71–80, 2006.

[2] B. Batagelj. Recognition of human faces using a hybrid system, PhD dissertation. Faculty of Computer and Information Science, University of Ljubljana, 2007.

[3] S. Gutta, J. Huang, P. Jonathon, and H. Wechsler. Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Trans. on Neural Networks*, 11(4):948–960, 2000.

[4] Infinitus. Digital video advertising. http://www.infinitus.si/, 2008.

[5] B. Lucas, T. Kanade. An iterative image registration technique with an application to stereo vision. In: Proc. 7th IJCAI, pages 674-679, 1981.

[6] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1424–1445, 2000.

[7] F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovač. Color-based face detection in the “15 seconds of fame” art installation. In *Mirage 2003, Conference on Computer Vision / Computer Graphics Collaboration for Model-based Imaging, Rendering, image Analysis and Graphical special Effects*, INRIA Rocquencourt, France, pages 38–47, 2003.

[8] F. Solina. 15 seconds of fame. *Leonardo*, 37(2):105–110, 125, 2004.

[9] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.