Interacting with Computers Advance Access published February 6, 2013

doi:10.1093/iwc/iws023

Audience Measurement of Digital Signage: Quantitative Study in Real-World Environment Using Computer Vision

Robert Ravnik and Franc Solina*

Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia
**Corresponding author: franc.solina@fri.uni-lj.si*

We present a quantitative study of digital signage audience measurement using computer vision. We developed a camera-enhanced digital signage display that acquires audience measurement metrics with computer vision algorithms. Temporal metrics of a person's dwell time, display in-view time and attention time are extracted. The system also determines demographic metrics of the gender and age group. The digital signage display was deployed in a real-world environment of a clothing boutique, where demographic and viewership data of 1294 store customers were recorded, manually verified and analysed. The analysis shows that 35% of customers specifically *looked-at* the display, having the average attention time of 0.7 s. Interestingly, the attention time was substantially higher for men (1.2 s) than for women (0.4 s). Age group comparison reveals that children (1–14 years) are the most responsive to the digital signage. Finally, the analysis shows that the average attention time is significantly higher when displaying the dynamic content (0.9 s) when compared with the static content (0.6 s).

Keywords: digital signage; audience measurement; quantitative study; computer vision

Editorial Board Member: Ruven Brooks

Received 1 May 2012; Revised 3 September 2012; Accepted 12 November 2012

1. INTRODUCTION

Modern applications of digital signage are interfaces to public or internal information, advertising, brand building and making enhanced customer experience (Krumm, 2011; Lundström, 2008; Müller *et al.*, 2011a; Schaeffler, 2008). Digital signage displays have the advantage over static signs because they can display the multimedia content such as images, animations, video and audio. The content can be adapted in real time to a different context and audience (Bauer and Spiekermann, 2011), making it attractive for use at airports, hotels, universities, retail stores and various outdoor public spaces. However, the displayed content is frequently generic and uninteresting for observers causing the effect of *Display Blindness* (Müller *et al.*, 2009). To make digital signage more effective as an information interface, the displayed content should be informative, dynamic and attractive.

The actual attention that people pay to public displays is one of the key parameters of digital signage. The comparative case study of Huang *et al.* (2008) reveals that paying attention to public displays is a complex process, which depends on several criteria such as positioning of the display, display size, content format and content dynamics. Therefore, to maximize the attention to digital signage, these parameters should be considered already during the design phase of the digital signage system. Research in digital signage today is aimed at exploring designs and options for delivering *engaging and interactive* content in public places. Michelis and Müller (2011) introduced *Audience Funnel*, a framework for audience interaction generalization. Various interaction modalities are proposed, including body position, speech, facial expression, body posture, gaze and touch (Müller *et al.*, 2010). Chen *et al.* (2009) describe a prototype system for interaction with digital

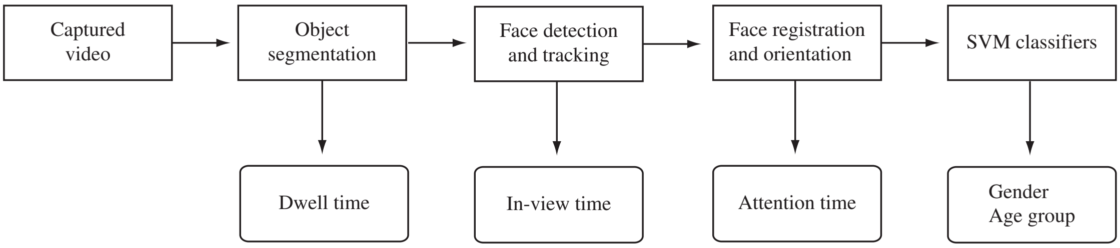

Figure 1. Scheme of computer vision enhanced digital signage system.

signage using hand gestures. Also, adaptive and interactive digital signage is permeating urban life and architecture (Kuikkaniemi *et al.*, 2011) as well as ubiquitous computing (Krumm, 2011). However, ubiquitous monitoring (Moran and Nakata, 2010) can lead to negative responses. Little and Briggs (2009) address the problem of receiving personal information in public spaces via personalized interaction. Grobelny and Michalski (2011) focus on digital signage content design. By performing pairwise human preferences comparison analysis, they show that different layout structures, background and content positioning can significantly affect the user's perception of the digital signage content.

Digital signage can yield a remarkable impact in commerce. A generalization study by Burke (2006, 2009) reveals that in-store digital signage increases customer traffic and sales. Indeed, the shoppers are the most responsive to messages that relate to the task at hand and their immediate interest. A qualitative study using questionnaires by Dennis *et al.* (2010) shows that digital signage is an effective stimulus, adding to positive perceptions of the mall environment, the emotions and the approach behaviour. Finally, digital signage screens also improve the image of shopping malls and create a favourable shopping atmosphere (Newman *et al.*, 2010).

Digital signage, clearly, is a strong contributor in various processes and fields; however, most of the measured impact was observed using qualitative methods. Nearly, all reported studies collected data by using interviews and questionnaires. Some of these publications address the inherent limitation of such approach and propose further research to determine whether people's actual behaviour really reflects their stated beliefs (Müller *et al.*, 2009; Newman *et al.*, 2010). Therefore, envisaging a *quantitative* method, with means for determining various audience measurement metrics could open a completely new window to digital signage, allowing for maximum interaction and continuous context awareness.

In this paper, we present a computer vision-enhanced digital signage system for monitoring the actual activity of the audience in front of the system and collect quantitative data on the audience, that is demographic metrics of a person's dwell time, display in-view time, attention time, gender and age group. The use of this system enables a new methodological approach to audience measurement in front of digital displays which gives quantitative data on all observed customers. We performed a quantitative field study in a real-world environment of a clothing boutique to test the methodology as well as the performance of the developed system. The collected data were then used for audience analysis of customers. The outline of the paper is as follows: Section 2 presents the audience measurement metrics and the computer vision-enhanced digital signage system, Section 3 elaborates the audience measurement field study, Section 4 presents experimental results, Section 5 provides a further analysis and discussion of the results and Section 6 gives final conclusions.

2. COMPUTER VISION-ENHANCED DIGITAL SIGNAGE

A real-time audience measurement system was developed for application in digital signage. The system is based on computer vision methods for detecting and tracking persons' faces from video. The video is captured by a digital camera that accompanies the digital signage screen. From the video, the system automatically computes various metrics and generates quantitative statistics of detected persons. The following temporal and demographic audience measurement metrics are determined: (i) dwell time which represents the sum of all time intervals when an observer was present in the same room or area as the display, (ii) in-view time which represents the duration of all time intervals when an observer was facing the display screen (without necessarily paying attention to the screen), (iii) attention time which is a part of the in-view time when an observer is actually looking at the display and (iv) the gender and age group which are demographic characteristics of each individual customer.

Our digital signage system consists of four video analysis modules, each designed for the determination of one of the metrics. Figure 1 illustrates the scheme of video analysis modules that are described in more detail below.

2.1. Dwell time

Object segmentation is used to determine the dwell time of each observer that enters the store. We employ a background

Figure 2. Field study. (a) Typical camera image from the clothing store where we performed the field study. (b) Customer observing the shop apparel. (c) Customer watching the screen. (d) Image of observer after object segmentation using background subtraction.

subtraction algorithm to extract foreground regions of the captured image and define potential presence of observers. Since the camera is static, we use a Mixture of Gaussians-based background modelling (Bouwmans *et al.*, 2008). Each image pixel is characterized by its intensity value in RGB space. The results of typical foreground subtraction are illustrated in Fig. 2d.

The segmented regions are tracked using the Fast Match Template algorithm supplied in OpenCV library (Bradski and Kaehler, 2008). This template matching algorithm is adapted for real-time video processing. The upper body part of an observer is used as a template image. Based on the comparison of module's results with the results of human annotators, this module is estimated to have an error of 10%.

2.2. In-view time

A frontal face detection algorithm is used to determine whether observers are facing the display. We use the Viola and Jones (2004) frontal face detector that runs in real time. The hit rate of this face detection method is reported to be 98% (Lienhart *et al.*, 2003), which is suitable for our purposes in terms of detection accuracy and speed. Using this face detector, we get the location of all present faces regardless of their position and scale down to the size of 20×20 pixels.

2.3. Attention time

The orientation of the observer's head is the central parameter in the determination of the attention time (when the observer is actually looking at the display). We use the multi-view active appearance model (AAM) method to register all detected faces. The AAM simultaneously models the intrinsic variation in shape and texture of the deformable visual object, a human face in this case, as a linear combination of basis modes of variation (Matthews and Baker, 2004). Although linear in both shape and appearance, overall, AAMs are non-linear parametric models in terms of the pixel intensities. Fitting an AAM to an image consists of minimizing the error between the input image and the closest model instance; i.e. solving a non-linear optimization problem. The reported convergence rate of this method is 98% (Saragih and Göcke, 2009). Using multi-view AAM registration and estimated observer's 3D position, we determine the observer's head orientation and consequently, if the head is oriented towards the display, this denotes the person's attention.

2.4. Gender and age group classifiers

The demographic metric of age and gender is determined within seven age groups: 1–14, 15–24, 25–34, 35–44, 45–54, 55–64 and over 65 years, all either male or female. We apply the support vector machine (SVM) learning algorithm for the age and gender classification (Moghaddam and Yang, 2002). The FERET database (Phillips *et al.*, 2000) was used as a learning set for gender and age classifiers. The FERET database comes fully annotated including facial images and the corresponding gender and year of birth data for 856 individuals. We use the AAM facial registration method described in Section 2.3 to register a face and warp it to the normalized frontal form of size 50×50 pixels. Normalized FERET faces were used to train SVM classifiers for gender and age. Using this approach, we achieved 91% classification accuracy on the FERET testing set.

3. FIELD STUDY

A field study of the proposed digital signage system was performed in a real-world environment, specifically focusing on

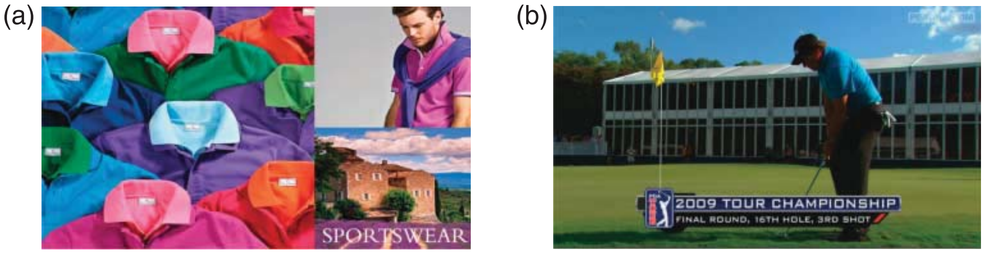

Figure 3. Broadcasting content of the digital signage display during the field study. (a) Static content type. (b) Dynamic content type.

the attention of the observers, i.e. their time metrics. We used a 24-inch Sony Vaio VPCL135FX/B computer display enhanced with a Logitech WebCam Pro 9000 camera. The digital signage system was positioned into a small clothing boutique in the city center of Ljubljana, capital of Slovenia, with a population of 300 000. The floor plan consisted of a main area ($\sim$35 m^2) situated between the entrance and the cashier's desk (see Fig. 2a) with an additional room in the back used for changing.

We have selected for the field study on purpose a small retail shop so that the entire retail space could be covered by a single-camera unit. We should mention that the shop sells higher priced sports fashion clothing and apparel, which can affect the demographic and behaviour characteristics of the customers.

To achieve the highest attention rates of our signage display, we optimized different criteria according to Müller *et al.* (2009) and Huang *et al.* (2008). To optimize the position, the display was mounted at eye level on a special shelf next to the cashier's desk, facing directly the entrance. For the eye-catching criterion, the shelves immediately next to the display were filled with small textile goods that were of immediate eye-catching interest. To obtain data for assessment of the animated content criterion, the static and dynamic content was displayed at the signage display during the field study. The static content consisted of a slide show with 20 slides shown in 10 s intervals. The slides showed pictures of distinctive sportsmen and sportswomen wearing attire from the shop's assortment (Fig. 3a). The dynamic content consisted of three video clips, which showed various sports and entertainment situations (Fig. 3b). The slide show and the videos were designed also to maximize the colourful content criterion, the emotional content criterion and the aesthetic look criterion.

3.1. Privacy aspects

Privacy-by-design (Brey, 2005; Langheinrich, 2001) as well as privacy-by-architecture (Spiekermann and Cranor, 2009) principles is incorporated in our computer vision-enhanced digital signage architecture, to ensure secure and appropriate handling with the acquired personal data. By design, all image processing is performed by the display unit in real time, therefore no visual records are stored or distributed over network. The display unit discards video image immediately after processing, storing only audience measurement metrics that are sent to the central server using encrypted data transfer. Using the proposed approach, we can acquire relevant audience data and perform a generalized behavioural analysis without the need to single out or even identify individual customers which is typically needed for a qualitative analysis using interviews or questionnaires.

Although the system processed video data in real time during the field study, video was recorded solely for the purpose of data verification by human reviewers. The manual annotation was performed within 3 days, discarding video afterwards. All customers in the shop were notified of the video recording, in compliance with the national privacy legislation.

3.2. Verification of results

The field study was performed within 23 daily sessions, consisting of a totally 214 h of video recordings. The tested digital signage system acquired characteristics and attention responses of 1294 people. To ensure the ecological validity of collected data, all automatically obtained data (temporal and demographic metrics) were manually verified by two human reviewers. We devised a video annotation programme for manual processing, following guidelines for effective video annotation proposed by Chen *et al.* (2008). Cohen's kappa coefficient κ was used for the evaluation of the inter-rater agreement (Carletta, 1996) of demographic metrics. Based on the data collected in this study, we determined $\kappa_{\text{gender}} = 1.0$ for gender classification and $\kappa_{\text{age\_group}} = 0.91$ for the estimation of observers' age groups.

The accuracy of automatically obtained parameters was compared with annotated data. The comparison shows that the system performs with a high accuracy, giving gender classifier 86.6% and age classifier 77.1% classification accuracy. The performance benchmark shows that the system is capable of video processing at 21 FPS using two cores of Intel Q8400 (2.66 Ghz) processor making it suitable for broadcasting adaptive content in real time.

We performed also a Kruskal–Wallis (K–W) test to determine the statistical significance of specific audience measurement metrics (Spurrier, 2003). The K–W test is a non-parametric method for testing whether measured data originate from the

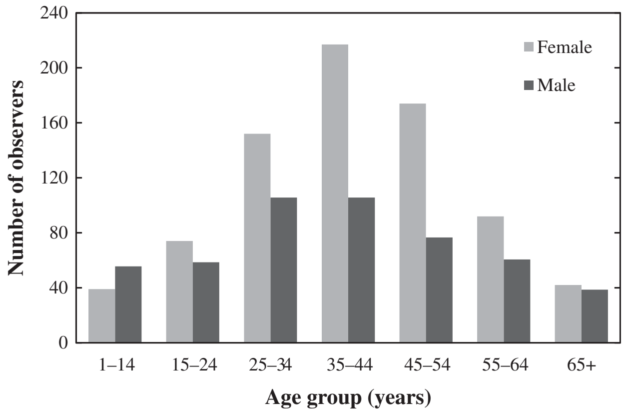

Figure 4. Distribution of observers according to age and gender in the field study.

same distribution. This method was chosen since it covers general (not necessarily normal) distributions, as observed in our extracted data.

Using manually verified data obtained from gender and age classifiers presented in Section 2.4, the study reveals that 61% of the acquired sample of customers were female and 39% were male. The age distribution was as follows: 7% in 1–14 years, 10% in 15–24 years, 20% in 25–34 years, 25% in 35–44 years, 19% in 45–54 years, 12% in 55–64 years and 7% in 65+ years age group. The full presentation of age and gender structure of the acquired sample is presented in Fig. 4.

Finally, we would like to comment that, in the pre-processing phase, all retail personnel audience data were manually excluded. In addition to the original data, we identified and excluded also 12 outliers, i.e. people whose dwell, in-view or attention time was 30 times over the mean.

4. RESULTS

The full results of the analysis are presented in Table 1. Note, that the table summarizes the results of three general tests performed for the audience metrics of gender, age group and content. Further, for dwell, in-view and attention time, the columns present: the number of analysed observers/customers (N), average dwell/in-view/attention time (mean), median (median) and standard deviation of mean (SD). Results of the two-tailed K–W test ($\alpha = 0.05$) are presented with: mean rank, test result value (H), degrees of freedom (DF) and the representative P-value.

Table 1 shows a large standard deviation in all three time metrics: dwell, in-view and attention time, which interestingly implies strongly varying behaviour of shop customers. Indeed, some people stayed in the shop for <20 s, whereas others were there for over half an hour, which expectably results in high standard deviation.

We next present individually each of the three time metrics.

4.1. Dwell time

The overall mean of dwell time, the time when a person is in the same room as the display, is 144 s (see Table 1, row 14, column 4). On average, each observer re-entered the scene 1.8 times. More specifically, the distribution of dwell times for all observers is presented in Fig. 5.

Comparison of the mean dwell time for gender reveals that male shoppers have a higher mean dwell time (156 s) than women (137 s) (see Table 1, rows 3 and 4, column 4). Also, the K–W test confirms the significant difference in mean ranks distribution ($H(1) = 4.25$, $P = 0.039$). Age comparison shows that the age group of 15–24 years has the mean dwell time substantially below average (101 s when compared with average 144 s). The difference in distribution is also confirmed using the K–W test ($H(6) = 20.4$, $P = 0.002$). Indeed, this quantitatively confirms that the boutique aims at an older target age group, between 25 and 55 years; which is also evident from Fig. 4. According to the mean comparison and the K–W test ($H(1) = 1.48$, $P = 0.223$), content type has no significant effect on the dwell time.

Interpreting the results, we could reason that the observed difference in the distribution of dwell time between males and females is due to the difference in the number of short shopping visits. Indeed, there are 51% of all females and only 44% of all males that have dwell time <60 s.

4.2. In-view time

In-view time analysis shows that the display comes into the field of view of an average person 4.9 times. The corresponding average of total in-view time is 17.6 s (see Table 1, row 27, column 4), indicating that the average person (customer) was facing the display for 12% of the total (dwell) time when the person spent in the room with the display. Distribution of the in-view time is presented in Fig. 6.

Gender comparison reveals higher in-view time for males. A significant difference in distributions is also confirmed by the K–W test ($H(1) = 32.4$, $P \leq 0.0001$). No significant effect on the in-view time is found for the metrics of age ($H(6) = 6.77$, $P = 0.343$) and displayed content ($H(1) = 0.85$, $P = 0.357$).

4.3. Attention time

The analysis reveals that 35% of all people entering the store looked at the display at least once, 12% looked at the display at least twice and 6% three times or more. The corresponding total average attention time of an average person was 0.7 s (see Table 1, row 40, column 4). The conversion rate between people engaging with the display and the ones not paying any attention to it at all was 35%, which relates well with the conversion rate of 33% reported by Michelis and Müller (2011). Distribution of attention time is presented in Fig. 7.

Table 1. Quantitative results of the digital signage audience measurement field study

Var.	Value	*N*	Mean	Median	SD	Mean rank	*H*	DF	*P*
Dwell time									
Gender	Male	504	**156**	73.4	204	674.3	**4.25**	1	**0.039**
	Female	790	**137**	60.3	193	630.4			
Age group	1–14	95	148	60.1	186	631.3	**20.4**	6	**0.002**
	15–24	133	**101**	43.7	146	521.4			
	25–34	258	154	64.4	222	650.6			
	35–44	323	138	68.7	191	648.9			
	45–54	251	163	86.8	206	687.6			
	55–64	153	157	72.1	213	691.8			
	65+	81	124	67.6	158	650.4			
Content	Slides	665	141	58.1	193	635.1	1.48	1	0.223
	Video	629	148	72.3	202	660.5			
Overall		1294	**144**	64.8	198				
In-view time									
Gender	Male	504	**20.9**	10.4	27.7	721.5	**32.4**	1	<**0.0001**
	Female	709	**15.6**	6.88	22.6	600.3			
Age group	1–14	95	17.9	8.15	26.7	654.7	6.77	6	0.343
	15–24	133	14.5	6.15	19.2	595.4			
	25–34	258	18.9	8.87	27.1	662.1			
	35–44	323	16.2	7.95	24.1	629.9			
	45–54	251	18.7	8.20	26.1	642.8			
	55–64	153	20.4	10.4	26.3	697.7			
	65+	81	15.7	8.55	18.3	667.8			
Content	Slides	665	16.7	8.1	22.7	637.6	0.85	1	0.357
	Video	629	18.6	8.8	26.7	656.8			
Overall		1294	**17.6**	8.38	24.8				
Attention time									
Gender	Male	504	**1.19**	0.0	2.61	741.7	**71.9**	1	<**0.0001**
	Female	790	**0.42**	0.0	1.19	587.4			
Age group	1–14	95	**2.39**	0.55	4.54	**815.1**	**37.6**	6	<**0.0001**
	15–24	133	0.70	0.0	1.41	663.6			
	25–34	258	0.60	0.0	1.35	638.2			
	35–44	323	0.42	0.0	1.19	589.7			
	45–54	251	0.67	0.0	1.69	647.5			
	55–64	153	0.68	0.0	1.73	659.9			
	65+	81	0.66	0.0	1.29	660.9			
Content	Slides	665	**0.60**	0.0	1.49	625.8	**5.71**	1	**0.017**
	Video	629	**0.86**	0.0	2.27	670.4			
Overall		1294	**0.72**	0.0	1.91				

Values of mean, median and standard deviation are given in seconds.

Interestingly, males are more attracted to digital signage than females: 48% of all males and only 27% of all females looked at the display at least once. The overall average attention time for males was 1.2 s and for females 0.4 s (see Table 1, rows 29 and 30, column 4). Significant difference in distribution was also confirmed using K–W analysis ($H(1) = 71.9$, $P \leq 0.0001$).

The age group shows a strong impact on the attention time. The K–W test shows a significant difference in distributions ($H(6) = 37.6$, $P \leq 0.0001$). Observing evident difference in the mean attention time for the 1–14 age group (see Table 1, row 31, column 4), we performed a two-tailed Steel–Dwass–Critchlow–Fligner multiple pairwise comparison *post hoc* test (Hollander and Wolfe, 1999) which confirms statistically significant difference between the 1–14 group and all other age groups. We believe that the reason for the youngest age group being so distinctive is in shop goods. Retail assortment offered

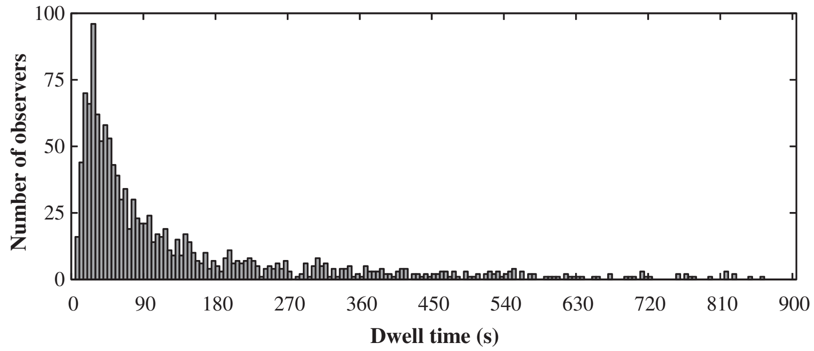

Figure 5. Distribution of dwell times for all observers.

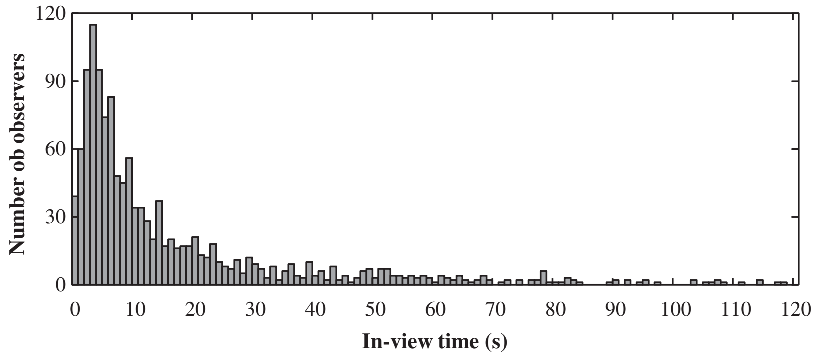

Figure 6. Distribution of in-view time for all observers.

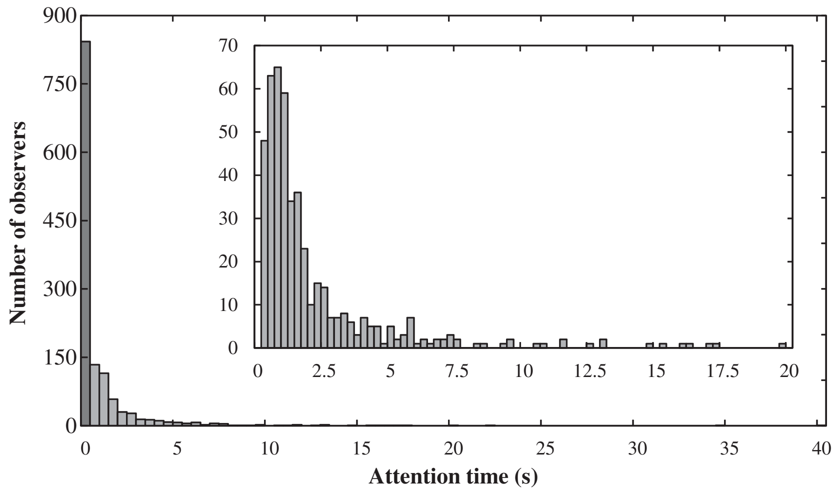

Figure 7. Distribution of attention time. The outer chart shows the distribution of overall attention time for all observers. The dark column represents the percentage of people that did not look at display at all (zero attention time). The inner chart illustrates the distribution of attention time for observers who looked at the display at least once.

nearly only adult apparel and children therefore directed their attention rather to the digital display.

Content type has no significant effect neither on dwell time nor on in-view time; however, it has an effect on the attention time (see Table 1, rows 38 and 39). The evaluation confirms that the dynamic content draws ∼1.5 times more attention than the static content. More specifically, the average attention time increased for 43% when broadcasting dynamic content. The results agree well with the qualitative digital signage observations (Dennis *et al.*, 2010; Huang *et al.*, 2008; Müller *et al.*, 2009) as well as with psychological studies on attention capture (Hillstrom and Yantis, 1994; Remington *et al.*, 1992). Statistical significance was also confirmed using the K–W test ($H(1) = 5.71$, $P = 0.017$).

4.4. Summary of analysed metrics by gender, age and content type

Gender: gender has a significant impact on all three observed temporal metrics. Men are more receptive for digital signage than women, having on average higher dwell time, in-view time and attention time (see Table 1).

Age: age has no effect on in-view time; however, it affects dwell time and attention time. Children (1–14 years) demonstrate the highest attention time, whereas the age group of 35–44 shows the lowest attention time.

Content: content (static or dynamic) does not affect the dwell and in-view time. However, broadcasting dynamic content shows a strong increase (43%) in attention time.

5. DISCUSSION

5.1. Interaction graph analysis

To better understand how the studied metrics are interrelated, we performed an interaction graph analysis (Jakulin and Bratko, 2004). Interaction graphs are based on entropy, which is a measure of the uncertainty in information theory. Each node in the graph corresponds to one of the observed metrics. The information gain of each metric is expressed as a percentage of eliminated uncertainty, written below the metric's name. Edges represent interaction between nodes as a value of their relative mutual information. A negative interaction edge, presented with an undirected dashed line, implies that the two metrics provide partly the same information. A positive interaction edge, presented with solid bidirectional arrows, indicates the amount of novel information added by the pair of connected metrics (Jakulin and Bratko, 2004).

The interaction graph of the data acquired in our field study was calculated using the Orange machine learning framework (Curk *et al.*, 2005) and is illustrated in Fig. 8. Attention time (as a continuous variable) was replaced with a binary class variable *looked-at* that indicates whether one has looked-at the display or not. The continuous metrics (dwell time and in-view time) were discretized using entropy MLD discretization. To ensure better legibility of the interaction graph, only the most significant edges of each metric are shown.

The most important metric is *in-view time* that alone eliminates 28.7% of uncertainty whether one looked at the display or not. The second most important metric is *in-view num*, indicating the number of times one had the display in view, that

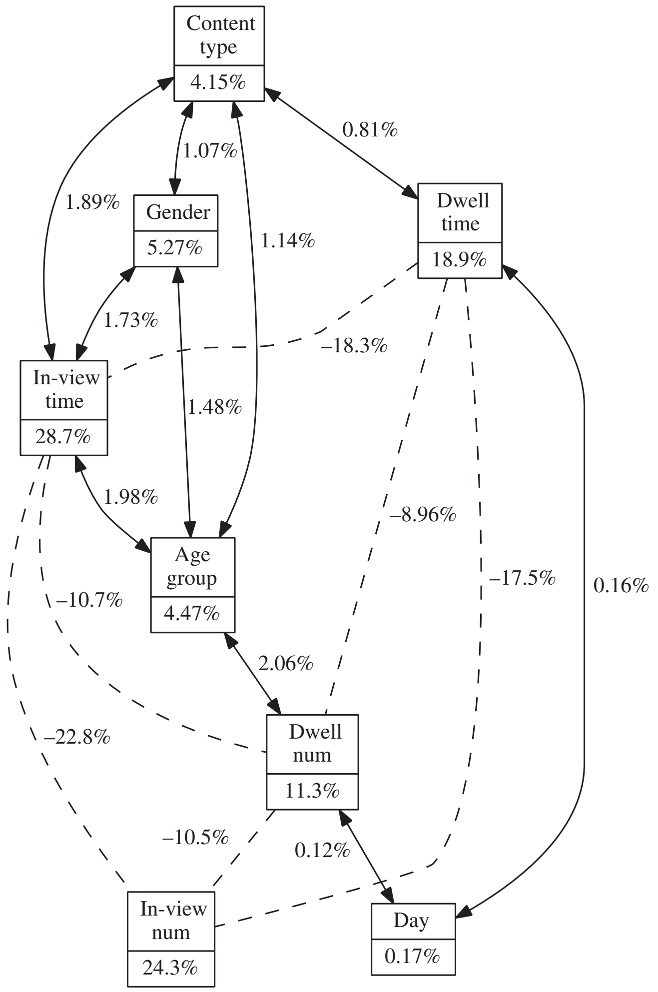

Figure 8. Interaction graph of metrics data collected in the field study.

alone removes 24.3% of class entropy. A negative interaction edge between these two metrics (dashed line) indicates that *in-view num* reduces class entropy by only $24.3 - 22.8 = 1.5\%$ on its own, once we have already accounted for *in-view time*. *Gender* alone provides 5.27% of information, but if we account for the positive interaction between *in-view time* and *gender* (solid bidirectional arrow), they together eliminate $28.7 + 5.27 + 1.73 = 35.7\%$ of class entropy.

The most informative metrics are: *in-view time* (28.7%), *in-view num* (24.3%), *dwell time* (18.9%) and *dwell num* (11.3%). We explain this by reasoning that the longer person stays in shop or the longer the display is in his/hers field of view, the higher is the probability of observing the screen. There is also a strong negative interaction between these metrics indicating significant amount of shared mutual information between them.

Metrics *gender* (5.27%), *age group* (4.47%) and *content type* (4.15%) are also all informative. Positive interactions between them and positive information gain imply that they are inter-dependent. Thus by knowing *gender* and *content type*, we eliminate 10.5% of class uncertainty.

The remaining metric *day* indicates the day of the week when the observer visited the shop. This node was added to the interaction graph to illustrate the validity of the proposed analysis. We can strongly believe that the day of the week does not correlate with the fact whether a shopper did look at the screen or not. The interaction graph analysis confirms this assumption since by knowing the day of the week, we eliminate only 0.17% of class entropy.

5.2. Limitations and benefits of the proposed digital signage system

There are several limitations but also benefits that one should consider when evaluating the presented results. The proposed methodology gives us detailed and measurable data on the behaviour of *all* observed subjects. Therefore, the number of customers analysed in such quantitative fashion is much larger than the number of customers typically covered within the framework of a qualitative analysis using interviews or questionnaires. It does not, however, offer an explanation of their behaviour. To fully explain and understand the social aspects of diverse attention rates, a quantitative audience measurement field study should be performed in parallel with collection of qualitative data from the same population. Having access to exact behavioural quantitative data on attention, one can improve also the preparation of questionnaires and interviews for qualitative studies.

An interesting result in our experiment which asks for an explanation is that males are more attracted to digital signage than females (Section 4.3). This finding could be explained, for example, using interviews or questionnaires. We could also test a hypothesis with the help of our quantitative methodology. For example, the observed result in our field study that men were more attracted to digital signage could be a consequence of the topic of the displayed dynamic content, mainly sports in our case. To test this hypothesis, we could replace the sport topics with, for example, family activities, to see if the corresponding attention data would change. We believe, however, that in order to explain such observations, the analysis of customer behaviour should go even deeper. A customer must first select an item, then she or he can make a decision to buy it and, finally, by paying for the item ending the whole purchase process. When a group of customers comes in a shop these roles for selecting, buying and paying can be distributed among different people in the group. The explanation for the above observation could therefore be that, in the framework of this field study, women were on the average more involved in the selection of clothing goods than men. Men who were part of a group of customers could be, in the meantime, more responsive to the whole store environment, including digital signage. Since the selection phase in the whole purchasing process typically takes longer than all other purchasing phases, this could also explain the collected statistics of attention time. By analogy, in a more technically oriented store, we would probably observe a reversal of gender statistics.

The broadcasting of information by means of the digital signage system in the presented field study was nearly optimal due to the highly controlled environment. The digital signage display and the complete broadcasting area were indoors, ensuring constant lightning conditions. Since computer vision is sensitive to changes in illumination, audience measurement errors would increase in an outdoor environment. The display was positioned on a prominent location at eye-level height where the camera was able to cover almost the entire shop area. All these and other parameters can hardly be optimal in an arbitrary real-world situation and can adversely affect the results of the proposed audience measurement system.

The performance of the proposed system also depends on technical characteristics. Real-time video processing requires a certain amount of processing power. The proposed system utilizes two cores of Intel Q8400 processor which means that the system is suitable for implementation on embedded devices. By using more complex computer vision algorithms or more expensive hardware components, such as infrared sensors or multiple cameras for stereo vision, the system's accuracy could be improved. However, we believe that the proposed set of computer vision methods offers a good price–performance ratio and can operate in real time on low-priced hardware. Selected computer vision algorithms are also copyright free, which makes the proposed system setup suitable for commercial deployment in large numbers.

It is probably already within technical means that camera-equipped digital signage systems could remember and identify individual customers by their appearance. But this would require establishing customer databases which would run against most privacy regulations. Although most customers are quite willing to identify themselves through various loyalty cards when they actually make a purchase, identifying themselves by merely stepping into a store is a completely different issue.

6. CONCLUSION

An advanced digital signage system is presented, consisting of a display screen, a digital camera and audience measurement software. Computer vision and machine learning methods were implemented for an automatic assessment of the audience measurement time and demographic metrics. The field study that we performed shows that, if the environment is optimally controlled, computer vision is at a stage where it can give fully reliable data for audience measurement research.

The digital signage system was applied in a field study for customer research in a clothing boutique, enabling a full quantitative audience measurement research. The attention time quantifier reveals that, on average, men pay attention to the digital signage display for 1.2 s, whereas women only 0.4 s. Age group comparison shows that attention time to digital signage is the highest (2.4 s) in the children age group (1–14 years) when compared with the all average attention time of 0.7 s. Interestingly, the average attention time is the lowest in the 35–44 years age group (0.42 s). The contents quantifier, dynamic or static, shows that broadcasting dynamic and not static digital signage content increases attention time for 43%.

More generally, these results are aimed to improve the future design of digital signage systems. The proposed architecture can be implemented on new or existing digital signage systems. Providing proper privacy care, it could serve as an advanced quantitative tool for various types of audience measurements. Additional social implications could be inferred and a wider behavioural analysis could be performed based on this non-invasive approach. We believe that using the proposed methodology, detailed audience measurement analysis can be performed continuously on all observed customers and at large number of points in parallel which could offer new ways of customer communication in marketing and advertisement. In this way, future digital signage systems could automatically adapt themselves to different types of retailers, to different customer bases and to different localities using the type of analysis of collected measurements which was used in this article only *post hoc*. Since demographic metrics are obtained in real time, they could be used to adapt content also in near real time to reflect actual variations in customers during a single day. Attention behaviour analysis and the possibility of designing adaptive scenarios could lead to a further evolution in content development for digital signage systems. Therefore, we plan to study in the future what kind of feedback mechanism would be the most efficient and simple enough for adaptive digital signage systems. Finally, as the next steps towards maximum-impact digital signage, the role of the display position, the display size and the design of the adaptive content itself should be studied carefully in the future. Additional machine learning algorithms could be applied on collected statistics to reveal significant customer behavioural patterns.

We plan to research in the future also different phases and roles in the purchasing process. A single customer must first select and decide to buy before making the purchase. If several people are in a group that entered the store, these roles are usually distributed among them. Can we determine automatically from video who is the initiator, decider and purchaser in a group that has made a purchase? How does the digital signage system influence each phase or role in the purchasing process? Is there a correspondence between a person's role in the purchasing process and its attention to digital signage? Additional studies should explore how the interaction design, usage of adaptive content and of different modalities for interaction affect the observer behaviour and attention towards digital signage. A comparative field study, using a quantitative audience measurement system in combination with a qualitative approach would offer additional insights on the user's perception of digital signage and his actual behaviour.

ACKNOWLEDGEMENTS

This work was supported by the Slovenian Research Agency, research program Computer Vision (P2-0214). We acknowledge the cooperation and support of the boutique store in performing the field study.

REFERENCES

Bauer, C. and Spiekermann, S. (2011) Conceptualizing Context for Pervasive Advertising. In Müller *et al.* (2011b), pp. 159–183.

Bouwmans, T., Baf, F.E. and Vachon, B. (2008) Background modeling using mixture of Gaussians for foreground detection—a survey. Recent Patents Comput. Sci., 1, 219–237.

Bradski, G. and Kaehler, A. (2008) Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media.

Brey, P. (2005) Freedom and privacy in ambient intelligence. Ethics Inf. Technol., 7, 157–166.

Burke, R.R. (2006) The Third Wave of Marketing Intelligence. In Krafft, M. and Mantrala, M. (eds) Retailing in the 21st Century: Current and Future Trends, pp. 159–171. Springer.

Burke, R.R. (2009) Behavioral effects of digital signage. J. Advert. Res., 49, 180–185.

Carletta, J. (1996) Assessing agreement on classification tasks: the kappa statistic. Comput. Linguist., 22, 249–254.

Chen, L., Chen, G., Xu, C., March, J. and Benford, S. (2008) Emoplayer: a media player for video clips with affective annotations. Interact. Comput., 20, 17–28.

Chen, Q., Malric, F., Zhang, Y., Abid, M., Cordeiro, A., Petriu, E.M. and Georganas, N.D. (2009) Interacting with digital signage using hand gestures. In Kamel, M.S. and Campilho, A.C. (eds) ICIAR, pp. 347–358. Springer.

Curk, T., Demsar, J., Xu, Q., Leban, G., Petrovic, U., Bratko, I., Shaulsky, G. and Zupan, B. (2005) Microarray data mining with visual programming. Bioinformatics, 21, 396–398.

Dennis, C., Newman, A., Michon, R., Brakus, J.J. and Wright, L.T. (2010) The mediating effects of perception and emotion: digital signage in mall atmospherics. J. Retail. Consum. Serv., 17, 205–215.

Grobelny, J. and Michalski, R. (2011) Various approaches to a human preference analysis in a digital signage display design. Hum. Fact. Ergon. Manuf., 21, 529–542.

Hillstrom, A.P. and Yantis, S. (1994) Visual motion and attentional capture. Percept. Psychophys., 55, 399–411.

Hollander, M. and Wolfe, D.A. (1999) Nonparametric Statistical Methods. Wiley, New York.

Huang, E.M., Koster, A. and Borchers, J. (2008) Overcoming Assumptions and Uncovering Practices: When Does the Public

Really Look at Public Displays? In Indulska, J., Patterson, D.J., Rodden, T. and Ott, M. (eds) Pervasive, pp. 228–243. Springer.

Jakulin, A. and Bratko, I. (2004) Testing the Significance of Attribute Interactions. In Brodley, C.E. (ed.), ICML, ACM. pp. 409–416.

Krumm, J. (2011) Ubiquitous advertising: the killer application for the 21st century. IEEE Pervasive Comput., 10, 66–73.

Kuikkaniemi, K., Jacucci, G., Turpeinen, M., Hoggan, E.E. and Müller, J. (2011) From space to stage: how interactive screens will change urban life. IEEE Comput., 44, 40–47.

Langheinrich, M. (2001) Privacy by Design—Principles of Privacy-Aware Ubiquitous Systems. In Gregory D. Abowd, Barry Brumitt and Steven A. Shafer (eds) Ubicomp 2001: Ubiquitous Computing, Third International Conference Atlanta, Georgia, USA, September 30 - October 2, 2001, Lecture Notes in Computer Science, pp. 273–291. Springer, London, UK.

Lienhart, R., Kuranov, A. and Pisarevsky, V. (2003) Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. In Bernd Michaelis and Gerald Krell (eds), DAGM-Symposium, Pattern Recognition, 25th DAGM Symposium, Magdeburg, Germany, September 10–12, 2003, Lecture Notes in Computer Science, pp. 297–304. Springer, Magdeburg, Germany.

Little, L. and Briggs, P. (2009) Private whispers/public eyes: is receiving highly personal information in a public place stressful? Interact. Comput., 21, 316–322.

Lundström, L.I. (2008) Digital Signage Broadcasting: Content Management and Distribution Techniques. Focal Press, Oxford, GB.

Matthews, I. and Baker, S. (2004) Active appearance models revisited. Int. J. Comput. Vis., 60, 135–164.

Michelis, D. and Müller, J. (2011) The audience funnel: observations of gesture based interaction with multiple large displays in a city center. Int. J. Hum. Comput. Interact., 27, 562–579.

Moghaddam, B. and Yang, M.H. (2002) Learning gender with support faces. IEEE Trans. Pattern Anal. Mach. Intell., 24, 707–711.

Moran, S. and Nakata, K. (2010) Ubiquitous monitoring and user behaviour: a preliminary model. JAISE, 2, 67–80.

Müller, H.J., Alt, F. and Michelis, D. (2011a) Pervasive Advertising. In Müller *et al.* (2011b), pp. 1–29.

Müller, H.J., Alt, F. and Michelis, D. (eds) (2011b) Pervasive Advertising. Human–Computer Interaction Series. Springer.

Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T. and Krüger, A. (2009) Display Blindness: The Effect of Expectations on Attention Towards Digital Signage. In Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B. and Tobe, Y. (eds) Pervasive, pp. 1–8. Springer.

Müller, J., Alt, F., Michelis, D. and Schmidt, A. (2010) Requirements and Design Space for Interactive Public Sisplays. In Bimbo, A.D., Chang, S.F. and Smeulders, A.W.M. (eds) ACM Multimedia, ACM, pp. 1285–1294.

Newman, A., Dennis, C., Wright, L.T. and King, T. (2010) Shoppers' experiences of digital signage-a cross-national qualitative study. JDCTA, 4, 50–57.

Phillips, J., Moon, H., Rizvi, S.A. and Rauss, P.J., (2000) The feret evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell., 22, 1090–1104.

Remington, R.W., Johnston, J.C. and Yantis, S. (1992) Involuntary attentional capture by abrupt onsets. Percept. Psychophys., 51, 279–290.

Saragih, J. and Göcke, R. (2009) Learning aam fitting through simulation. Pattern Recognit., 42, 2628–2636.

Schaeffler, J. (2008) Digital Signage: Software, Networks, Advertising, and Displays: A Primer for Understanding the Business. Focal Press.

Spiekermann, S. and Cranor, L.F. (2009) Engineering privacy. IEEE Trans. Software Eng., 35, 67–82.

Spurrier, J.D. (2003) On the null distribution of the Kruskal–Wallis statistic. J. Nonparametr. Stat., 15, 685–691.

Viola, P.A. and Jones, M.J. (2004) Robust real-time face detection. Int. J. Comput. Vis., 57, 137–154.