Oriented trajectories as a method for audience measurement

Manuel López-Palma*, Josep Ramon Morros†, Javier Gago‡ and Montserrat Corbalán‡
*Venco Electrónica, L'Hospitalet de Llobregat, Spain. mlopezpalma@mlopezpalma.com
†Signal Theory and Communications Dept., Universitat Politècnica de Catalunya, Barcelona, Spain
‡Electronic Engineering Dept., Universitat Politècnica de Catalunya, 08222 Terrassa, Spain

***Abstract*—The quantification of the attention received by advertisements is of paramount importance to determine its effectiveness. In this work, a simple and effective objective method for the assessment of the attention given to advertisements is provided. The proposed method is based on computing the oriented trajectory of the different test persons along with their head pose. This way, it is possible to determine if a given person is looking towards the advertisement. While other similar methods use more complex setups, requiring a camera at each advertisement location, our method needs only a single (or a few) ceiling camera. Even though the apparent simplicity, the method can compute attention measures at each point of the scene.**

I. Introduction

Several methods have been established to determine the degree of attention received by an advertisement. The goal is the objective comparison of the attention taken by different ads, either static or dynamic (screens). The objective is twofold; on the one hand, to determine the effectiveness of the ad, determined either by its content or by its place of visualization; on the other hand, to be able to more accurately charge for the advertising content.

To be useful, the determination of the attention metrics must be robust, non-invasive and adaptable to different environments, often cluttered with many objects. Another important requirement is that the complete setup has to be as cheap as possible to make it competitive in a variety of situations.

There are some metrics commonly used to evaluate audience measurement [1]: some of the most usual are:

- *Dwell time (DT)*: The total amount of time a observer spends in the same area as the sign being evaluated.
- *In-view time (IVT)*: The total amount of time that the observer is facing the sign (not necessarily paying attention to the screen).
- *Attention/Engagement time (AT)*: The total amount of time the observer is actively looking at the sign. The Attention time allows to quantify the degree of attention a given sign has received.

In this contest, we have that necessarily $DT > IVT > AT$.

Most methods to determine these quantities require a camera at each analyzed point [2]. A camera placed over the ad facing the customer can measure whether a customer is actively looking towards the ad or not. The drawback of these methods is that they require a camera for each measurement point, which can be cumbersome and affected by occlusions.

We propose a method that, relying only in top-view cameras, can determine audience measurements in all points of the room. Top-view cameras are non-intrusive, cost effective, are almost immune to occlusion problems between costumers and can alleviate privacy problems. The cameras are used to determine the trajectory of the persons and their head orientations. With these measures and by taking into account physiological parameters of the human vision, we can estimate the values for the Dwell, In-view and Attention times at all points of the room Moreover, additional information can be obtaine d, such as the distance of viewing and the relative angle of viewing, thus allowing a richer analysis of the scene. In this sense, our method can be seen as a generalization of the methods previously described, as it allows determining metrics at any point and also allows to capture more information. The method has been designed for static (analog or digital) signage although it can be easily extended to dynamic digital signage.

As an example of the type of analysis that our method permits, we will introduce a new measure, the focus of attention received at each point of the room. This measure quantifies the amount of attention a region receives during a period of time. While the temporal metrics are useful for defined targets (signs on the walls, for instance), this new measure allows to find the room spots that receive more attention. We will justify in Sect. V that the attention to a target depends on the distance so, for instance, to be able to see the same details in an image at double distance this image has to be twice as large. The new measure will be based not only on the amount of time that a target is observed but also in the distance to the observer plus other parameters that affect the attention (speed of movement, angle of vision with respect to the trajectory, etc.). This technique can provide values for all the objects in a room, not only of those located at the walls. This allows to determine the objects in a store that are more attractive to users.

The main contributions of this paper are:

- A non-invasive, cost-effective method to compute the temporal metrics used to quantify the attention received by a sign.
- A method to evaluate the intensity of focus of attention at all points of the room.
- A system to determine the regions that receive more attention in a given room.
- To test the method

This paper is organized as follows: Sect. II provides a review of the state-of-the-art of audience measurement and related technologies. In Sect. III, a more detailed view at the audience metrics is given. The proposed system is detailed in Sect. IV and Sect. V. Experimental validation of the proposed system is given in Sect. VI. Finally, conclusions are drawn in Sect. VII.

II. Related work

In the literature there is a good amount of works using computer vision and other sensing technologies to analyze the attention that people pay to public signs. A good review can be found at [2]. An important factor in these methods is the placement of the cameras. The majority of methods use frontal-view or top-view cameras. Frontal cameras can be situated at or near the sign and take a frontal view of the customer. The advantages of this setup are that from a frontal position the face and eyes of the customer can be detected, making possible a fine analysis of the direction of the gaze. Additionally, information such as the identity, age or gender of the customers can be extracted. However, this setup requires a camera at each analysis position and is affected by occlusions and has privacy considerations. Another popular setup is to use top-view ceiling cameras. Top-view camera-based tracking is a non-invasive method to estimate the trajectories of people in indoor environments and can avoid the privacy and occlusion problems suffered by front-facing cameras if multiple users interact with each other. A single camera can analyze several spots resulting in cost-effective solutions. The drawbacks are that the system can not capture the face/eyes of the customers, so information such as age and gender can not be determined.

In [1], a study of digital signage audience measurement using cameras located at the signage displays and facing the customers is presented. Temporal metrics of a persons Dwell time, display In-view time and Attention time are extracted by body and face detection and head pose estimation. An estimation of the gender and age of the customers are also obtained. A new approach to automatic modelling of a retail store consumer behaviour based on audience measurement data is introduced in [3]. Among other parameters, the In-view and Attention time have been used. They show that under controlled environment the viewership data can be used to predict purchase decisions.

Eye-tracking technology has been used to analyze the direction of the customer gaze and to determine if he/she is actively looking at a given sign or product. For instance, [4] investigate the visual saliency of in-store signage and products and how this saliency affects to the customer decisions. The analysis is done by using data from eye-track and sales data from grocery stores. Eye-tracking is also used in [5] to investigate the role and limitations of peripheral vision for preference-based choice tasks in a real supermarket setting.

RGB-D cameras are a choice in many works [6], [7], [8], [9] because the ability to capture depth information additionally to RGB significantly simplifies segmentation of the persons' bodies and limbs, allowing a more precise and powerful analysis. In [6], in addition to a sparse array of

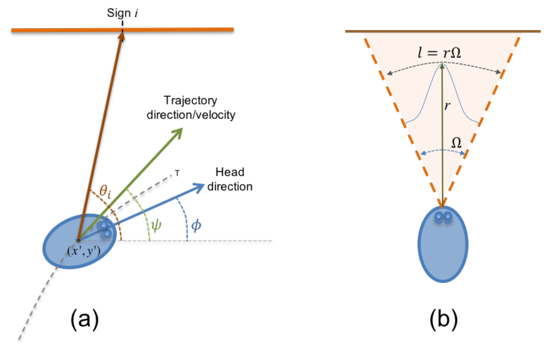


Fig. 1: a) State variables b) Angle of vision

RGB-D cameras, other sensor modalities (active radio beacons emitted by customers mobile devices) are used to determine the customer localization in the store. This allows to capture information of the consumer activities in the store.

A popular use of RGB-D sensors is to place them as top-view cameras. For instance, in [7] a RGB-D camera is located above a shelf. Several customer behaviors are analyzed, including the reaching gesture, browsing and weighing up a book, etc.. Their method estimates which shelf the customer reaches as well as the gesture itself. A related approach is the one in [8] where a real-time human posture and activity recognition system is proposed, with a single top-view depth-sensing camera. This method is capable of tracking users' positions and orientations, as well as recognizing postures and activities (standing, sitting, pointing, etc.).

III. Oriented trajectories

Our primary analysis is based on the computation of the Dwell, In-view and Attention times from the oriented trajectories. The oriented trajectories are computed by determining, at each frame captured from a top-view camera, the position (x', y') and head orientation of the user, relative to the room coordinates. For simplicity, we assume a rectangular room and the camera to be aligned with the room.

A person inside the room can be parametrized using an state vector $\mathbf{x}$:

$$\mathbf{x} = [p', \psi, v, \phi] \tag{1}$$

where $p' = (x', y')$ indicates the position of the person in the room coordinate system, ψ is the direction defined by the person's trajectory, v is the instantaneous velocity of the person and ϕ is the angle of the head, also in the room coordinate system (see Fig. 1 (a)).

An oriented trajectory is defined as the temporal sequence of states for all the times instants k a person is in the field of view of the camera: $\mathbf{T} = \{\mathbf{x}_k\}$.

In order to be able to compute the temporal information (Attention time and In-view time), for each of the objects of interest (signs) we also capture at each time instant the angle that forms the line connecting the center of the head and the center of the sign, θ_k (see Fig. 1 (a)).

IV. Temporal analysis

The determination of the temporal metrics (Attention, In-view and Dwell time) are based solely on the angles θ_k, ϕ_k captured at each time instant. The Dwell time is simply the total amount of time that the customer is visible by the top-view camera. In-view time can be defined as the time the user is able to see the sign. There are several definitions of the In-view time. For instance, the Media Rating Council's' (MRC) guideline for an In-view ad [10] is when more than 50% of the ad is in view for more than 1 second. On the other side, [1] defines the In-view time as the case when a camera located at the top of the ad can detect the face of the customer using a frontal-face detector. We will use an approach similar to [1] as it can be related with the measured angles θ, ϕ. In fact, a typical frontal face detector such as the one in OpenCV [11], [12] can detect slightly non-frontal faces, up to angles or approximately 45°in both directions. This is equivalent to consider that a customer has an ad in-view when $|\theta - \phi| \leq 45°$ (See Fig. 2). To avoid spurious detections we will keep the requirement stated in [10] that the sign has to be in-view for more than one second. For this, we will analyze the temporal sequence $\{\mathbf{x}_k\}$, classifying each instant as in-view ($\mathbf{x}_k^{iv}$) or not-in-view. We extract the subsequences where all consecutive angles are classified as in-view and that are larger than one-second $\{\mathbf{x}_k^{iv}\}_j$ (j denoting sub-sequence index). Then for a given trajectory i, the In-view time is:

$$T_i^{iv} = \sum_j length(\{\mathbf{x}_k^{iv}\}_{i,j}) \cdot t_f \tag{2}$$

where t_f is the duration of the instant (video frame) that depends only on the video frame rate.

For the computation of the Attention time, the requirement is that the customer is actively looking at the ad. In [1], this was determined by estimating the gaze direction of the customer using an AAM model of the face. Our equivalent approach is to consider the cases where $|\theta - \phi|$ is small enough so the ad is inside the cone of vision where the customer is capable of full attention, this is, when $|\theta - \phi| \leq 25°$. We have

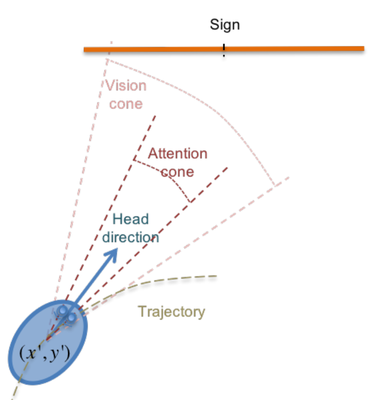


Fig. 2: Representation of the In-view and Attention vision angles

derived this value based on studies of the physiology of the visual field [13]. As previously, we classify each instant as attention ($\mathbf{x}_k^a$) or non-attention and we keep the subsequences of consecutive attention instants $\{\mathbf{x}_k^a\}_j$ that are larger than one second. For a given trajectory, the Attention time is:

$$T_i^a = \sum_j length(\{\mathbf{x}_k^a\}_{i,j}) \cdot t_f \tag{3}$$

By summing over all the trajectories, the final values of the Attention (T^a) and In-view (T^{iv}) times can be computed:

$$T^a = \sum_i T_i^a, \quad T^{iv} = \sum_i T_i^{iv} \tag{4}$$

V. Extended analysis

In this section, a generalization of the concept of Attention time is proposed. This metric can be computed for each point inside the room, thus providing an estimate of the probability of a group of persons to focus their attention on a given region or spot inside a room. This metric includes some important factors not considered in the attention time, such as the distance from the customer to the evaluated point, the trajectory velocity or the difference between trajectory direction and head orientation.

The method is based on determining the trajectories and directions of visualizations of all the individuals entering the visualization zone. This determination of the direction of visualization is performed by measuring the head's orientation in each instant of time. This is, we approximate the gaze direction by using the direction of the head and no determination of the direction of the eyes is performed, as it would require a complex and expensive multi-camera configuration.

A. Instantaneous attention

Based on the previous considerations, an instantaneous estimation of the probability at each location can be obtained. Then, this probability will be integrated for the duration of the trajectory of each individual and averaged for the different individuals during the evaluation.

For each located head and in each time instant, the visualization zone is determined by an angular sector (of angular span Ω) that goes from the head's location until the limit of the room in the direction of the head (See Fig. 1 (b)). We consider that all points $p = (x, y)$ inside this visualization zone to receive an increase in the received attention that can be explained by a 'visual ray' from p' to p.

Inside this angular sector, the attention of a person in any arc at distance r is considered constant. Let $A(r)$ be the amount of attention over this arc and $l = r\Omega$ the angular span of the arc. Then, this property can be expressed by:

$$A(r) \cdot l = C_0 \tag{5}$$

being C_0 a constant. Thus, if Ω is fixed, $A(r) = C_1/r$. We will assume that constant C_1 is the same for all the different persons. C_1 can be determined by normalizing the probability maps at the last step of the process.

We consider the amount of attention to be maximal in the direction of the head (ϕ) and to decay exponentially as we look to a point at an angle α from this direction:

$$A(\alpha) = A_0 \exp\left\{ - \frac{(|\alpha - \phi|)}{2\sigma^2} \right\} \tag{6}$$

where A_0 represents the value of the attention at angle ϕ (the head's angle in the room coordinate system) and σ determines the velocity of the exponential decay.

In addition to the effects of distance and head's angle, we will consider also the velocity of the person and the relative position of the head with respect to the person's trajectory. We consider that the degree of attention varies according to the walking speed as:

$$A(v) = \frac{1}{\kappa + v} \tag{7}$$

where κ is a small regularization constant and v is given by the difference of the positions of the head in successive frames.

When persons walk, they usually look into the direction of their trajectory (represented by ψ). Deviations of the gaze direction from the trajectory direction indicate a strong interest in some object along this direction. The effect of the angle of the head (ϕ) with respect to the trajectory direction (ψ) is modeled by a function depending on this angular difference:

$$A(|\psi - \phi|) = 1 + C_2|\psi - \phi| \tag{8}$$

The complete instantaneous attention function for a point $\mathbf{p} = (x, y)$ given that the person's head is located at $\mathbf{p}' = (x', y')$ and oriented along ϕ will be obtained as a product of all the partial attentions:

$$A(\mathbf{p}, X) = A(r) \cdot A(\alpha) \cdot A(|\psi - \phi|) \cdot A(v) \tag{9}$$

B. Single trajectory attention

Individuals will move inside the room from a starting point to an exit point. The trajectory followed by an individual i can be represented by the evolution of the state sequence at each discrete intervals k:

$$T_i = x_i^{0:k} = \{x_i^0, \cdots x_i^k\} \tag{10}$$

An example of a trajectory is presented in Fig. 6 (a). Trajectory direction ψ is represented using a white line. The red arrows indicate the instantaneous direction of te head ϕ and the yellow boxes mark some of the detections of the head.

The computation of the attention for a given trajectory consists of integrating the attention function A in (9) in the interval $0 : k$. As the time is discrete, the integration is in fact a summation.

$$A_i(\mathbf{p}) = \frac{\sum_k A_i^k(x, y, X_i^k)}{\sum_p \sum_k A_i(x, y, X_i^k)} \tag{11}$$

This attention function represents an indication of the normalized attention of an individual at a given point. This is, the likelihood of each point to be observed by the individual.

C. Multiple trajectory attention

To evaluate the attention provided by multiple trajectories (multiple individuals), the individual trajectories will be added and normalized.

$$A(\mathbf{p}) = \frac{\sum_{i=1}^{N} A_i(p)}{\sum_{i=1}^{N} \sum_p A^i(p)} \tag{12}$$

VI. Experimental results

A. Computation of times

To validate the computation of view and attention times, a set of recordings have been captured with a ceiling camera. In these recordings, several individuals walk in predetermined trajectories across the room, while looking at the four posters that are affixed to each of the four walls of the room.

The experimental set-up is illustrated in Fig. 3. The figure shows a top-view diagram of the room with four posters (red, orange, white and green) on each of the walls. Note that the fact that the walls are outside the field of view of the ceiling camera is not a problem for the proposed algorithm.

The picture at the center of the figure shows the view from the ceiling camera. The lines on the floor mark the trajectories the individuals are asked to follow. The posters are colored A4 paper, glued to the wall at a height of 150 cm from the floor.

There are two types of trajectories: linear and circular. In linear trajectories (marked on the floor by strips of tape of different colors), the individuals are instructed to look at only one poster for the complete trajectory. There are 4 stripes that can be walked in two directions and four posters, resulting in 32 combinations. For the circular trajectories, the individuals look consecutively at the nearest poster (a different one in each quadrant of the trajectory).

Each trajectory lasts 12.5 seconds or 250 frames (recorded at 20 Hz) in each trajectory. Individuals of different characteristics have been recorded: men and women, with hair and without hair, high and low, with a hat and without a hat.

For each recording, the position and the direction of the head at each frame is manually annotated, to create ground truth data of the trajectories. This will allow to validate the method by comparing the times automatically computed using the proposed algorithm and the ground truth data. As our goal

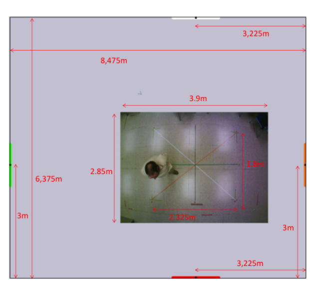


Fig. 3: Experimental setup

Fig. 4: Example of trajectories. Vectors at each point of the trajectory indicate the direction of the head

is to demonstrate the validity of the method to measure times, we have opted for manual annotation instead of using any of the existing tracking algorithms. For a real application, there are several tracking algorithms for top-view cameras (see for instance [14] for a review of methods) with excellent performance that can be used for this purpose. In particular, using a particle filter method it is possible to track both the position and head angle simultaneously. Fig. 4 shows the results of 2 trajectories (rectilinear and circular).

For each recorded trajectory, the manual trajectory annotations have been used to compute the angles of vision from each point of the trajectory to the corresponding posters.

Dwell, In-view, and Attention times have been calculated as in (4). The Dwell time is based on the number of frames in which the individual is completely detected. The In-view time is calculated as the number of consolidated frames at 45°. A frame is considered as consolidated at a given angle α if during the next 20 frames (1s) the viewing angle $|\theta - \phi|$ remains equal to or less than α. Attention time is calculated by using, a consolidation angle of 25°, as indicated in Section IV.

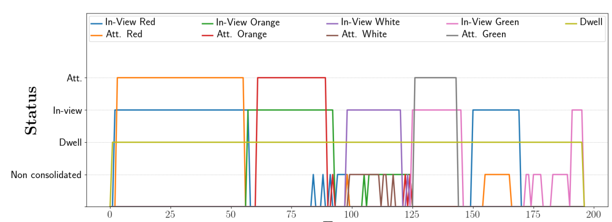


Fig. 5: Temporal analysis of the circular trajectory

The results are summarized in Table I and Fig. 5. For each trajectory and each target, the computed viewing times (*Dwell*, T_i^d / *In-view*, T_i^{vi} / *Attention*, T_i^a) are shown. Times are indicated in frames (fr) and their equivalent in seconds (s). In the linear trajectory, the test person looks at the red target during the whole recording. In the circular trajectory, the person looks alternately and consecutively at the four targets.

The times computed for the linear trajectory correspond completely with the indications given to the persons, showing that only the red sign is viewed. In this case, the three times are almost the same. In the circular trajectory, the In-view time is shared among the different targets. The red target has a higher viewing time because the person is walking more slowly in this part of the trajectory. The Attention times are lower than the In-view times, which is logical as the individual is slowly

TABLE I: Viewing times on a linear and circular path

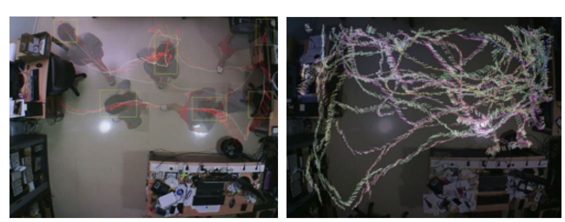
Trajectory	**Target**	**Values**		
		T_i^d	T_i^{iv}	T_i^a
Linear	Red	70 fr 3.50s	67 fr 3.35s	63 fr 3.15s
	Orange	70 fr 3.50s	0 fr 0.00s	0 fr 0.00s
	White	70 fr 3.50s	0 fr 0.00s	0 fr 0.00s
	Green	70 fr 3.50s	0 fr 0.00s	0 fr 0.00s
Circular	Red	195 fr 9.75s	76 fr 3.80s	53 fr 2.65s
	Orange	195 fr 9.75s	36 fr 1.80s	29 fr 1.45s
	White	195 fr 9.75s	23 fr 1.15s	0 fr 0.00s
	Green	195 fr 9.75s	26 fr 1.30s	20 fr 1.00s

turning from one target to the another. The Attention time for the white target is zero because the person has a tangential trajectory too close to the wall and he's never looking at the target for more than one second, so the attention is never consolidated. The other targets also have a small amount of non-consolidated frames (see Fig. 5). To further validate the method, 21 rectilinear trajectories and 20 different circular trajectories have been analyzed, obtaining in all of them results according to the instructions given to the test persons. This show that the proposed method allows to measure Dwell, In-view, and and Attention times in a simple way.

B. Computation of density of attention

The density of attention measure has been applied to a practical case in which N_S different persons walk a room and observe objects without any given guidelines. Fig. 6 shows an example of a trajectory and all the N_T trajectories in the test.

As in the previous case, each frame has been manually annotated with the position and head orientation of the persons. The annotations have been used to compute individual $A_i(\mathbf{p})$ (11) and total $A(\mathbf{p})$ (12) attention functions for all points in the room. The analysis is restricted to N_R regions of interest (ROIs), $r = 1 \cdots N_R$, as shown in Fig. 7 (in this case, $N_R = 3$). We compute the amount of attention of a single test person to a given ROI, A_i^r as the average of $A_i(p)$ over all points p of the ROI. Similarly, the total attention over the ROI A^r is computed by summing the attentions.



(a) Single trajectory (b) All trajectories

Fig. 6: Trajectories and head angles in a real scenario

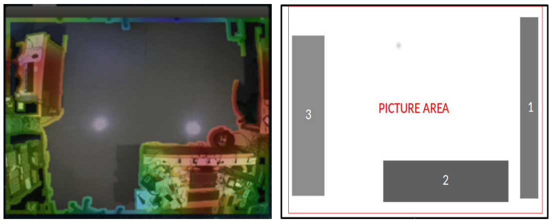


Fig. 7: Attention heath map and rectangular ROIs

The attention over a ROI, for one trajectory A_i^r and for all trajectories A^r, is computed as the average of $A_i(p)$ and $A(p)$ over all points p of the ROI.

To test the method, each person i assigns a value SV_i^r in the range [1-10] to each ROI r according to the interest paid to this ROI. An external observer also evaluates the interest OV_i^r that each ROI has raised in the person under test. The evaluations are averaged to compute the interest function I_i^r

$$I_i^r(\%) = \frac{SV_i^r + OV_i^r}{\sum_{k=1}^{N_R} SV_i^k + OV_i^k} \cdot 100 \tag{13}$$

The total interest over a region r is computed over all the trajectories:

$$I^r(\%) = \frac{\sum_{i=1}^{N_T} SV_i^r + OV_j^r}{\sum_{k=1}^{N_R} \sum_{i=1}^{N_T} SV_i^k + OV_i^k} \cdot 100 \tag{14}$$

The method is evaluated with the error between the interest I^r and the attention A^r computed using the proposed method.

$$E^r(\%) = \frac{|I^r - A^r|}{I^r} \tag{15}$$

To evaluate the method, we have recorded 4 persons walking a total of ten trajectories through the room (this is, $N_S = 4$ and $N_T = 10$) and looking at the three ROIs ($N_R = 3$). The results are presented in Table II.

TABLE II: Interests values and attention function

ROI	I^r (%)	A^r (%)	E^r (%)
1	38.97	38.58	1.01
2	28.72	31.57	9.03
3	32.31	29.85	8.24

Fig. 7 (left) shows a heat map showing in false color the amount of attention received over the ROIs. Errors are always below 10%, thus showing the ability of the method to determine the regions that receive more interest.

VII. Conclusions

The present article presents a novel technique for the measurement of human attention. This method is based in oriented trajectories captured using a top-view ceiling camera and presents several advantages over alternative methods of determining human attention in indoor environments: For one side, the use of a top view camera ensures that the method is cost-effective, non-intrusive, occlusion-free, and avoids privacy concerns. On the other side, it allows to obtain not only the metrics commonly used in determining the attention over specific spots (i.e. signs) but also allows an extended analysis of the customer behavior. The new proposed measures, also based on the oriented trajectories, allow computing the attention for each point in the room. This additional analysis would not be possible for systems based on front-view cameras. The experimental validation shows that this method can effectively be used to determine which areas received most attention or to compare the relative attention received by different objects. One possible drawback of the method is that it does not allow to discriminate the attention given to objects located at at a given place at different heights. In the future we plan to investigate the use of RGB+D sensors to solve this by estimating the pitch angle of the head. To analyze extended areas, an extension to multiple overlapping cameras should be investigated as well.

Acknowledgment

This work has been developed in the framework of projects TEC2016-75976-R, financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund (ERDF) and 2014-DI-067 financed by the Catalan government AGAUR.

References

[1] R. Ravnik and F. Solina, "Audience measurement of digital signage: Quantitative study in real-world environment using computer vision," *Interacting with Computers*, vol. 25, no. 3, pp. 218–228, 2013.

[2] M. Quintana, J. Menendez, F. Alvarez, and J. Lopez, "Improving retail efficiency through sensing technologies: A survey," *Pattern Recognition Letters*, vol. 81, pp. 3–10, 2016.

[3] R. Ravnik, F. Solina, and V. Zabkar, "Modelling in-store consumer behaviour using machine learning and digital signage audience measurement data," in *LNCS*, 2014.

[4] J. Clement, J. Aastrup, and S. Charlotte Forsberg, "Decisive visual saliency and consumers' in-store decisions," *Journal of Retailing and Consumer Services*, vol. 22, pp. 187–194, 2015.

[5] E. Wästlund, P. Shams, and T. Otterbring, "Unsold is unseen ... or is it? Examining the role of peripheral vision in the consumer choice process using eye-tracking methodology," *Appetite*, vol. 120, pp. 49–56, 2018.

[6] M. Sturari, D. Liciotti, R. Pierdicca, E. Frontoni, A. Mancini, M. Contigiani, and P. Zingaretti, "Robust and affordable retail customer profiling by vision and radio beacon sensor fusion," *Pattern Recognition Letters*, vol. 81, pp. 30–40, 2016.

[7] J. Yamamoto, K. Inoue, and M. Yoshioka, "Investigation of customer behavior analysis based on top-view depth camera," in *WACVW*, 2017.

[8] C.-J. Wu, S. Houben, and N. Marquardt, "EagleSense: Tracking People and Devices in Interactive Spaces using Real-Time Top-View Depth-Sensing," 2017, pp. 3929–3942.

[9] D. Liciotti, E. Frontoni, A. Mancini, and P. Zingaretti, "Pervasive system for consumer behaviour analysis in retail environments," in *FFER, VAAM*, 2016.

[10] VV.AA., "MRC viewable ad impression measurement guidelines," Media Rating Council, Technical Report, June 2014.

[11] P. Viola and M. Jones, "Robust real-time object detection," in *International Journal of Computer Vision*, vol. 57, no. 2, 2001, pp. 137–154.

[12] "OpenCV Libaries," http://opencv.org/.

[13] H. Traquair, *An Introduction to Clinical Perimetry*, 4th ed. London: Henry Kimpton, 1944.

[14] D. Liciotti, M. Paolanti, E. Frontoni, and P. Zingaretti, "People Detection and Tracking from an RGB-D Camera in Top-View Configuration: Review of Challenges and Applications," in *ICIAP*, 2017.