Who watches the watchers? Quality control of the human inspection in production lines using Visual Intensity of Attention

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*—On multiple occasions production lines require inspectors, human operators that visualize certain steps of the production and determine the quality of the resulting products. However, inspectors are subject to errors. We propose a method based on computer vision to decide if the inspector has used an adequate attention in the different points of inspection so that pieces that have not been verified can be marked for rejection or re-inspection. The method uses a top-view ceiling camera that computes the trajectories and areas of vision of the inspector, and determines which products have received the correct amount of attention. The resulting attention can be compared with the acceptance range in the inspection protocol to determine if the inspection is valid.**

I. Introduction

Product inspections are a key element of quality control that allow to verify the quality of the product on the site at different stages of the production process and prior to its dispatch. Inspection is an effective way to prevent quality problems. Nowadays, abundant automatic quality control methods relying on a variety of sensors are widely available. However, in many cases, human inspection is still needed. For instance, in factories where many products are produced, with short series, automatic inspection may not be practical. In other cases, the complexity of the inspection makes full automation very complicated or expensive. Finally, combining both automatic and human control can increase accuracy in critical scenarios. In all these cases, a human inspector is still essential and is the last barrier against the introduction of defects in the market. However, human inspectors are also prone to make mistakes. Conventional causes are fatigue, eye exhaustion, external distractions or other causes may result in reduced attention to the inspection task. In other cases, the operators of the production line may have a bonus that involves manufacturing a certain number of products during the day. If the production goes a little behind, a friendly inspector can give way to products without proper inspection to reach the bonus.

In this work, we propose a method of computing the attention that a subject gives to the different areas of a room a long time. We apply this method to determine if an inspector is paying the due attention in the verification of the points indicated in a certain inspection protocol. The method proposes is to ensure that the inspector performs the inspection protocol correctly, thus allowing to determine whether the faulty products were detectable according to the protocol or not. Another application is to determine the requirements (i.e. amount of sufficient inspection) to be applied based on inspection measures carried out with controlled inspectors.

The system is based on the determination of the visual intensity of attention by using oriented trajectories obtained with top-view ceiling cameras. Cameras in a top-view configuration are non-intrusive, cost effective, almost immune to occlusion problems between costumers and can alleviate privacy problems. The cameras are used to determine the trajectory of the inspectors and their head orientations. With these measures and by taking into account physiological parameters of the human vision, we can estimate a measure of the visual intensity of attention at different areas prefixed in advance. This measure quantifies the amount of attention a region receives during a period of time. This measure is 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 (angle of vision with respect to the trajectory, distance, etc.).

The contributions of this work are:

- A non-invasive, cost-effective method to evaluate the intensity of attention at all points of the room.
- A system to determine if an item in a production line has received a correct human inspection.

This paper is organized as follows: Sect. II provides a review of the state-of-the-art of audience measurement and related technologies. The proposed system to compute the visual intensity of attention is detailed in Sect. III. Sect. IV shows how this measure is applied for determining the quality of human inspection. Experimental validation of the proposed system is given in Sect. V. Finally, conclusions are provided in Sect. VI.

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 products in a space. Most of these studies have been realized to determine the attention given to products in a store or the attention paid to advertisements in order to quantify its effectiveness. Even if the goals of these works are different, the results are of interest in our application.

We will analyze two families of methods that use computer vision to perform this analysis: those using frontal-view cameras and those using a top-view configuration, as in our method.

Frontal cameras can be situated at or near the product or the inspection line and take a frontal view of the inspector. By using a frontal position, the face and eyes of the inspector can be detected, making possible a fine analysis of the direction of the gaze. On the negative side, this setup is usually more cumbersome, requiring a camera at each analysis position and being affected by occlusions. The top-view setup is non-invasive method and can avoid the occlusion problems of front-facing cameras. It is a more cost-effective solution as a single camera can analyze several inspection spots or even several inspectors, resulting in more cost-effective solutions. The drawbacks are the inability to capture the face/eyes of the inspectors, which results in an approximate and slightly less precise determination of the gaze direction.

Eye-tracking technology has been used to analyze the direction of persons' gaze and to determine if they are actively looking at a given spot. For instance, [1] 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 [2] 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 [3], [4], [5], [6] 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. A popular use of RGB-D sensors is to place them as top-view cameras. For instance, [5] proposes an human posture and activity recognition system using a top-view depth-sensing camera. This method is capable of tracking persons' positions and orientations, as well as recognizing postures and activities (standing, sitting, pointing, etc.).

III. Visual intensity of attention

The method is based on the concept of Intensity of Attention (IoA), a scalar quantity that, for a given point or region, quantifies the probability that an observer has focused his attention on this point. The IoA is influenced by factors such as the distance from the inspector to the evaluated point or the alignment of the inspected point and the gaze direction. Attention over regions or objects are obtained by summing the IoA values for all the points of the region.

The method is based on determining the oriented trajectories and directions of visualizations of the observer in the visualization zone. By measuring the person's head orientation at each instant of time, the direction of visualization (gaze) can be estimated.

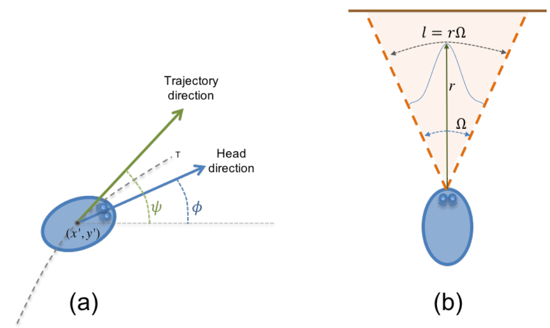


Fig. 1: a) State variables b) Angle of vision

A. Oriented trajectories

An oriented trajectory describes the evolution of an individual during the time of inspection. The scene is captured from a top-view camera. At each frame, the position (x', y') and head orientation of the user, relative to the room coordinates, is determined. Thus, a person inside the room can be parametrized using a state vector $\mathbf{x}$:

$$\mathbf{x} = [p', \psi, \phi] \quad (1)$$

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

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

B. Intensity of Attention computation

To compute the IoA, the trajectories and directions of visualizations of all the individuals entering the visualization zone are first determined by tracking the person's head position and its orientation angle at each time instant. Note that no determination of the direction of the eyes is performed, as eyes are not visible from the top-view camera. Eye tracking would require a more cumbersome frontal camera configuration.

C. Instantaneous attention

Once the oriented trajectories have been computed, the instantaneous attention at each location can be obtained. Then, this attention will be integrated for the duration of each trajectory and averaged for the different individuals during the evaluation.

For the 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:

$$A(r) \cdot l = C_0 \tag{2}$$

being C_0 a constant. Thus, if Ω is fixed, $A(r) = C_1/r$. Constant C_1 is assumed to be the same even for different persons. C_1 is determined by normalizing the attention maps at the last step of the process.

The amount of attention is considered 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:

$$A(\alpha) = A_0 \exp\left\{-\frac{(|\alpha - \phi|)}{2\sigma^2}\right\} \tag{3}$$

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.

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:

$$A(\mathbf{p}, X) = A(r) \cdot A(\alpha) \tag{4}$$

D. Trajectory attention

The trajectory followed by an individual i can be represented by the evolution of the state sequence at each discrete intervals k:

$$T_i = x_i^{0:k} = \{x_i^0, \cdots x_i^k\} \tag{5}$$

The computation of the intensity of attention for a given trajectory consists of integrating the attention function A in (4) in the interval $0 : k$. As the time is discrete, the integration is in fact a summation.

$$IoA_i(\mathbf{p}) = \frac{\sum\limits_k A_i^k(x, y, X_i^k)}{\sum\limits_p \sum\limits_k A_i(x, y, X_i^k)} \tag{6}$$

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.

IV. Measuring the quality of human inspection

A. IoA over a region

The method presented in the previous section allows to determine the IoA at any point viewed by the top-view camera. In order to quantify the amount of attention received by a given region (object) R_j, the average IoA over all points composing the object is computed:

$$IoA_i^j = \frac{\sum\limits_{p \in R_j} IoA_i(p)}{NP_{R_j}} \tag{7}$$

where NP_{R_j} is the number of pixels in region R_j.

The value of IoA measures the amount of attention the region or object receives. A threshold on this value can be used to determine if this attention was appropriate to correctly inspect the object. The threshold can be determined by calibration using measurements performed with controlled inspectors according to the inspection protocol.

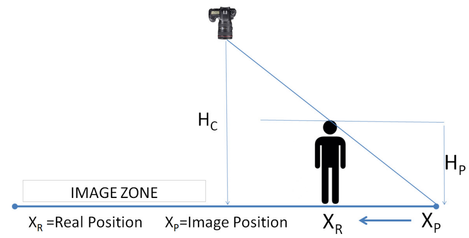


Fig. 2: Parallax correction

As commented before, the IoA takes into account not only the viewing time but also other important factors that may affect the quality of the inspection, such as the distance to the observed point or the observation angle.

B. Parallax correction

To obtain a precise estimation of the point at which the inspector is looking, the parallax error of both the person position and the inspected object position must be corrected. Fig. 2 depicts this process when correcting the parallax for the observer (the same approach is used for the inspected objects). In the top-view camera, the image of the person's head appears at the X_p position, however its real position is X_R. The correction can be performed using a simple trigonometric relationship:

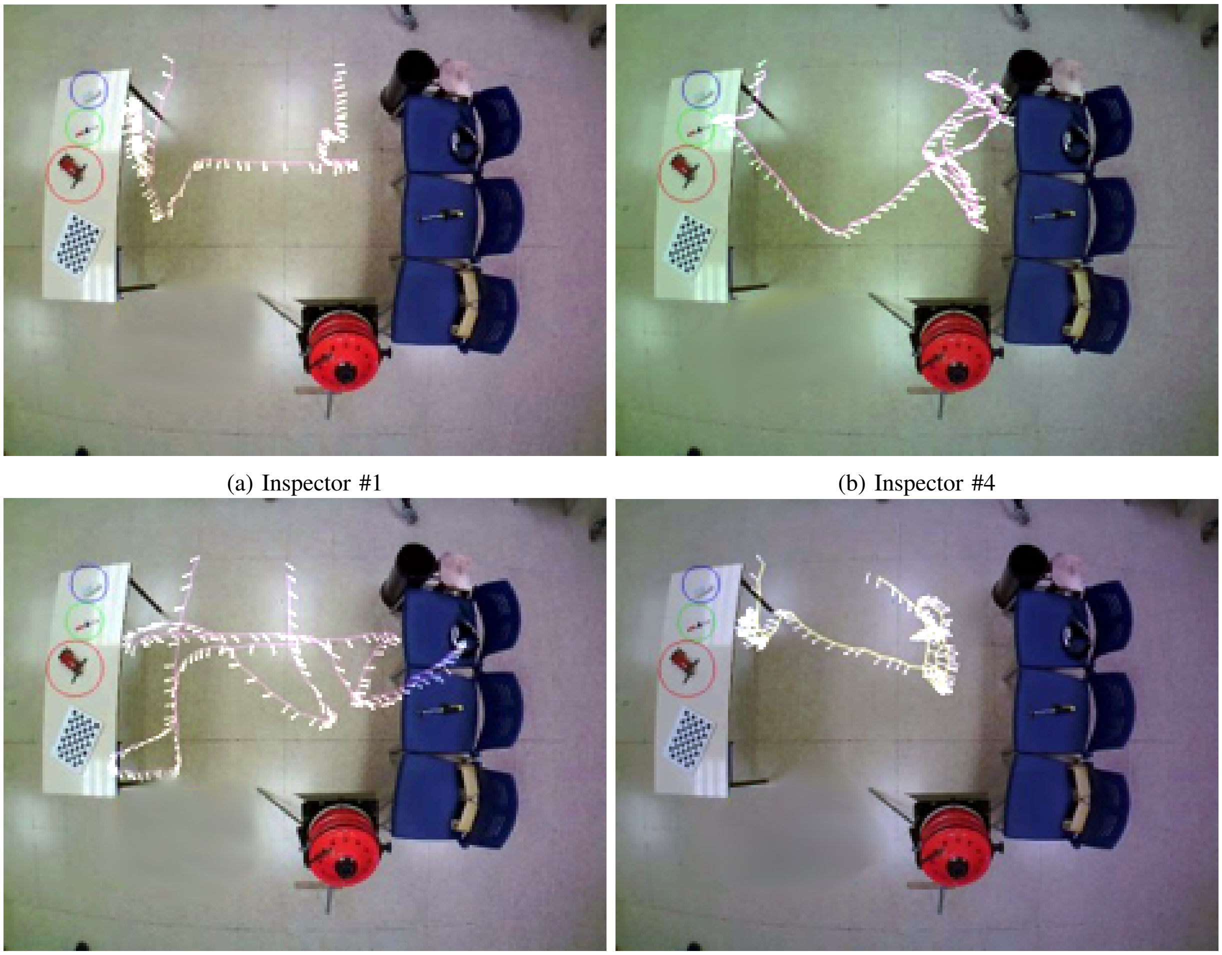
$$X_R = X_p \left(1 - \frac{H_p}{H_c}\right) \tag{8}$$

where H_p is the height of the person or object and H_c is the height of the camera. H_c depends on the camera setup, in our case, 350cm. For the inspected objects, the real value can be determined when deploying the system. In our tests we have used $H_p^{obj} = 70\,cm$. For persons, as their height can not be determined with precision in a real scenario, an estimation is used: $H_p^{pers} = 170\,cm$.

V. Experimental validation

The method has been applied to a a practical case in which different people walk freely, inspecting objects located in a given area. This setup simulates the inspection of objects in a production line. The goal is to determine if the inspectors have correctly examined the objects in Fig. 3, framed with a blue, green and red circles. To perform their function correctly, inspectors must pay a minimum of attention to each of the objects. The minimum degree of attention for each object depends on the specific application and must be determined in advance. The trajectories are rather arbitrary to clearly show that the method can analyze the inspection effectiveness even for complex trajectories.

We have analyzed 11 recordings where 11 inspectors perform a different trajectory each one. The method calculates the intensity of attention IoA_i^j of each inspector i and the three regions R_j, marked in the Fig.3. In addition to the IoA, the total amount of time (that is, the time T_i^j) each region R_j has



(a) Inspector #1 (b) Inspector #4

(c) Inspector #7 (d) Inspector #11

Fig. 3: Setup example with trajectories for four different inspectors. Images show the instantaneous position and head orientation of the inspectors.

been viewed by inspector i, is also counted for comparison purposes. Only the frames where the inspector is looking actively (that is, with an angle less than $15°$) at the target are considered, both for the computation of IoA and T_i^j.

The main difference between IoA_i^j and T_i^j is that, while T_i^j only considers the amount of time the object has been looked at, the IoA measure also reflects the effects of the viewing distance and the angle of vision.

The oriented trajectories have been manually annotated by marking the position and orientation of the head in the different frames of the recordings of the top-view camera. Note that our goal is to show the validity of the proposed method, so we have used a manual annotation instead of a tracking algorithm. There are several person tracking algorithms for top-view cameras with excellent performance (see for instance [7] for a review of methods) that could be used for this purpose. In a future work we will study the use of a particle filter tracker [8], which can simultaneously track the position and head angle of each person in the scene. Initial detection could be performed by using a foreground detection algorithm [9].

Videos have been recorded at 20 fps. The total number of frames in each trajectory i is indicated as T_i^T.

Table I shows the corresponding IoA_i^j results corresponding to the three blue, green and red regions (IoA_i^B, IoA_i^G, and IoA_i^R), the observation times T_i^B, T_i^G, and T_i^R frames, as well as the total time for each path, T_i^T.

In our example, we consider an object properly inspected if $IoA_i^j \geq 0.1$. This threshold is application dependent and is obtained by using the method with a reference inspector that follows strictly the inspection protocol. In the case of using the time-only method (T_i^j), we consider as valid inspections those performed during at least 0.5s at a close range.

Fig. 3 depicts the trajectories of 4 of these 11 inspectors, which will be analyzed in further detail. In these examples we study both the IoA and the T measures to quantify the attention of the inspectors in the corresponding zones.

Fig. 3 (a) corresponds to an inspector (#1) that has inspected the three zones correctly. This fact is properly reflected by the values of both IoA and T measures. In this case, both measures would provide a correct result.

Fig. 3 (b) corresponds to a case (inspector #4) where only one of the objects (the green one) has been properly inspected. Inspector #4 walks through the observation area but remains distant from the objects. For this reason, the values of the T measure reveal an intense inspection in the 3 areas of interest (0.5s, 0.5s, and 1.45s, respectively). However, the IoA measure discovers poor inspection in the blue ($IoA_4^B = 0.01$) and red ($IoA_4^R = 0.07$) zones, both less than the established 0.1 threshold for good inspection. This is because the inspection is mostly effectuated from far distances. Part of the trajectory consists of an approach to the objects' location but during this portion of the trajectory the inspector looks mainly at the blue object. In this case, IoA correctly decides that only the blue object is properly inspected, while T erroneously accepts inspections of objects red and green as correct.

Fig. 3 (c) shows an example (inspector #7) where an insufficient inspection has been applied to two of the three objects. The results of IoA reveal that inspection does not achieve the 0.1 threshold in the blue and green zones ($IoA_7^B = 0.002$, $IoA_7^G = 0.08$) and meets it by little in the red zone ($IoA_7^R = 0.11$), thus correctly classifying the three cases. However, the T measure incorrectly accepts the inspection of the green object as valid. Again the distance effect is decisive.

Finally, Fig. 3 (d) corresponds to a case (inspector#11) with good inspection of the blue and green zones and insufficient inspection for the red zone. Again, IoA values identify correctly these three cases, while the T measure would accept the three as valid, failing at the red zone because of remote observation.

The wrongly classified cases are shown in red in Table I. In total, 6 out of 33 cases (an 18%) would be accepted incorrectly as valid by the method using only the time of visualization (T).

VI. Conclusion

This article presents a new technique to evaluate the correct human inspection of different points or pieces based on computer vision. This method is based on the determination of the oriented trajectories captured from a top-view ceiling camera. The use of this type of cameras is cost-effective, non-intrusive, occlusion-free and avoids privacy concerns. For the evaluation of the inspection, two types of measures have been presented. The first one is based solely on the amount of time during which the inspector has examined the region of interest with an angle of viewing less than $15°$. This method does not take into account the distance of observation. For this reason, a new measure called Intensity of Attention is presented, which also takes into account the distance and the angle from which the region is inspected. This function provides consistent results throughout the series of tests performed and provides a much better determination of the quality of the inspection.

TABLE I: IoA and T values for the 11 trajectories. In red the cases where T wrongly classifies the validity of the inspection.

Inspector	$\mathbf{IoA^B}$	$\mathbf{IoA^G}$	$\mathbf{IoA^R}$	$\mathbf{T^B}$	$\mathbf{T^G}$	$\mathbf{T^R}$	$\mathbf{T^T}$
#1	0,33	0,66	0,78	0,90s	1,50s	1,75s	4,15s
#2	0,08	0,33	0,24	0,30s	1,40s	2,00s	3,70s
#3	0,10	0,43	0,37	0,50s	1,75s	1,50s	3,70s
#4	0,01	0,30	0,07	0,50s	0,50s	1,45s	2,40s
#5	0,05	0,10	0,14	0,70s	1,05s	1,90s	3,65s
#6	0,05	0,12	0,07	0,25s	0,90s	0,90s	2,00s
#7	0,00	0,08	0,11	0,05s	0,65s	1,05s	1,75s
#8	0,06	0,19	0,20	0,25s	0,70s	0,85s	1,80s
#9	0,12	0,12	0,11	0,95s	1,05s	1,15s	3,15s
#10	0,64	0,53	0,14	3,05s	3,10s	0,90s	7,05s
#11	0,60	0,31	0,09	1,70s	1,05s	0,60s	3,35s

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] 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.

[2] 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.

[3] 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.

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

[5] 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.

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

[7] 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.

[8] K. Nummiaro, E. Koller-Meier, and L. V. Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99 – 110, 2003.

[9] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1999.