# Visual-Attention-Driven Dynamically Changing Virtual Stores

Andrey Arguedas Espinoza
*Computer Science Engineering School*
*Costa Rica Institute of Technology*
San Jose, Costa Rica
and12rx12@gmail.com

***Abstract*—In this paper, we propose a system to create constantly changing virtual stores based on user's visual attention, search context and behavioral shopping predictability. Our main idea is to model visual attention by creating saliency maps of gaze vision in 3D Virtual Reality scenes in order to replace inventory based on user's attention and the real time search context. We implemented a model that constantly retrieves feedback from the user's visual attention with a Virtual Head Mounted device and starts to recommend new products and replace them dynamically into the virtual store with the help of a prioritization model based on the most gaze-attention received products, their matching characteristics and association rules with other products in a data set. The main objective of our proposal is to increase sales and improve Virtual Reality shopping experiences by showing user preferred related products understanding their real time visual attention data.**

***Index Terms*—Virtual Reality, Virtual Stores, Visual Attention, Gaze, Eye Tracking, Product Choice Prediction, Regression Model, Product Replacement, FOVE.**

## I. INTRODUCTION

Virtual Reality (VR) is becoming a more preferred option for shopping experience since it offers advantages over traditional or web-based shopping. VR stores increase positive emotions and perceived store attractiveness for users [1], this, combined with the costs of renting real state and maintenance fees increasing, might accelerate the aggregated value for retailers if they migrate their stores to the virtual world.

From the customer's perspective VR enables them to shop in different ways, being able to virtually try items and with the advantages of not need for transportation to the stores [2].

Taking in consideration the advantages of VR shopping, in this paper, we propose a dynamically changing VR store by designing a framework that takes eye tracking information from a VR device headset to detect user's preferences and interests in order to dynamically change the store inventory to a more user related one, therefore, enhancing shoppers' satisfaction and improving the engagement with brands and the retailers.

## II. RELATED WORK

It is important to start by researching the concepts that will allow us to implement a model like the one we proposed, so we focused our research in specific areas.

For our project we want to detect the Visual Attention of an user and drive this data into an improving shopping experience. Visual Attention is defined as set of mechanisms that limit some processing to a subset of incoming stimuli. This attentional mechanisms shape what we see, and allow for concurrent selection of relevant information and inhibition of other information [3].

In the Visual Attention Analysis we have found the work from [4], they used an eye-tracking VR headset to use visual attention data to arrange artwork in a virtual museum, posting banners for virtual events or placing advertisements by analyzing visual attention patterns of the users. Their propose consist in a data-driven optimization approach for automatically analyzing visual attention and placing visual elements in 3D virtual environments. With the collected eye-tracking data they train a regression model for predicting gaze duration, then use the predicted gaze duration output from the regressors to optimize the placement of visual elements with respect to certain visual attention.

Thanks to devices like the FOVE VR headset is easier to obtain this data in 3D VR environments [5]. It is important to understand that the allocation of visual attention is measured in terms of gaze duration on a given scene element (interval of viewing an element without shifting one's gaze) [6].

Thanks to previous studies [7] we know that there is a strong relationship between gaze signals and user interest, they proved this by using a built-in mobile camera to measure gaze duration while viewing areas of interest on a multi-column web page. Thanks to this knowledge we want to base our proposal.

The research of the visual attention is in a really good state, on 2013 researchers did an state of the art publication on numerous computational models devised for predicting the allocation of attention based on the distribution of visual features in a scene [8].

Our implementation bases in the process of analyze the allocation of attention in 3D scenes. Therefore, we have found these two projects that we think are good starting points to research on.

First, David et al [9] was able to compare four methods that can be used to generate 360 degrees saliency maps from eye tracking data obtained by a VR device. The studies of saliency maps is well established for the 2D world, where it is used to predict gaze [10], image or video compression [11] and others. Meanwhile in the VR world is different, the user is surrounded by a photorealistic virtual scene, due to the spherical nature of 360 degrees content, saliency map generation methods are different. Saliency maps aggregate data from multiple observers into a representative map of human attention. Typically, saliency maps are generated by summing fixations from each observer into a discrete fixation map. This fixation map is convolved with a 2D Gaussian kernel to produce a continuous map highlighting salient regions instead of individual pixels. We will implement their approach of a subsampled version of the modified Gaussian that proved to be sufficient for generating accurate 360 degrees saliency maps fairly quickly, even at image resolutions exceeding 8K [9].

Authors from [12] created a new visual attention user dataset for omnidirectional video (ODV) by investigating behavior of viewers when consuming ODV content. They found that the limitations of the ODV technologies are strongly related to the massive volume of video data that needs to be stored and rendered compared to traditional video. Since HMDs use only a fraction of an ODV at a time, namely viewport, ODV can be optimized by predicting where the viewers' visual attention is concentrated at a given point in time. For this, they used saliency maps, which predict viewer's eye fixations for given content. To understand the salient regions of ODV viewed in HMDs, saliency maps can be estimated either by collecting eye fixations during subjective tests or by using visual attention models. In their research they created a new dataset which include viewport trajectories (VT) and visual attention maps from 17 participants while watching uncompressed ODV, this can be used to obtain visual attention maps without the need for eye tracking devices. For modeling visual attention they were only interested in the user's fixations. They consider a valid fixation, if the Vieport Center Trajectory (VCT) remains almost stable in a certain location for at least 200 milliseconds. This requires clustering the VCT in order to remove influence from minor irrelevant movements and to reduce sensitivity to noise.

Now we know that is possible to track visual attention via different methods, therefore we want to know if other studies have use this type of data to predict future eye tracking or optimize different aspects of a 3D scene thanks to the obtained data. In this research path we have found the following contributions.

The authors from [13] were able to accommodate 2D items dynamically for more proactive visual exploration based on ongoing search context by analyzing the distribution of eye gaze through an eye-tracking device, in order to infer how the most attractive items lead to the finally wanted ones. To guide dynamic item arrangement, they employed an eye-tracker that takes spatiotemporal eye gaze distribution as input. They also construct a context map to retain the association rules among items by applying topic-based mining techniques to annotated texts associated with the items. This allow them to instantly infer the following favorite items from the most focused one by investigating its spatial neighbors on the context map. They also compared association rules among items obtained using our context map and those reproduced by conventional recommendation systems based on co-purchasing data.

Their solution was based in compute the spatiotemporal distribution of visual attention by convolving each gaze point in the sequence with a Gaussian weighting kernel. For a visualization technique they used the spatial distribution of visual attention as a heatmap, in which the color changes from blue to green to red as the degree of attention increases. One important aspect of their work four our project is that they increases the priority values of invisible items so that they could replace visible ones that are not of interest with this invisible ones. They also construct a context map that retains plausible association rules among items in diagram style. This means that the map represents pairwise relationships between the items as a spatial layout of the corresponding points on a 2D diagram. The context map provides them with association rules among the items so that they can select the items that will attract the viewers' attention next. This is similar to what we need for our VR store.

Another key aspect of our solution will be the predictability of possible preferred products by the user, on this topic we have learned that we need to understand buying behavior as a set of consumption habits that can be analyzed to help in predicting the needs of specific target audience [14]. With the data of consumption habits, we aim to achieve in increasing sales by generating better recommendations. Most of these product recommendations systems like [15] and [16] have tried different approaches of Machine Learning techniques to improve results, they have also done benchmarking between these approaches, gathering all their information we also implement a Random Forest Regressor Model since multiple trees reduce the likelihood of encountering a classifier that does not perform well due to the connection between the train and test data [14] and we also need to obtain multiple new recommendations.

We also want to highlight the contribution from [17], they created a system that worked as inspiration for our main idea. Their project consists in procedural generated interiors of buildings in 3D city scenes. Their solution creates rich furniture arrangements for all rooms of complex buildings and even for entire cities. The key idea is to only furnish the rooms in the vicinity of the viewer while the user explores a virtual building in real time. In order to compute the object layout they introduced an agent-based solution in charge of procedurally furnish a different 3D scene every time the user navigates a room. Our main idea is that our VR store behaves similarly constantly changing products based on user's search context and preferences.

We also took inspiration form the work of [18], were they were able to optimize product placement in Virtual Stores and increase exposure of products by a total cost function that encodes spatial and product exposure constraints taking in consideration aspects such as size, texture, height and others. Their approach was implemented with a regressor model for predicting product exposure and a Markov Chain Monte Carlo Algorithm to find optimal placement.

Taking in consideration all this related work we are proposing a solution that takes different techniques form this works and modifies them to create a model that automatically changes the visible inventory in a VR store based on the user's visual attention feedback and the predictability of the customer's preferences based on the attributes of the products that received more attention from the user.

This paper contributes to expand the range of solutions where we can use the previous techniques mentioned, it also deepens in the generation of saliency maps in VR environments and focuses in the understanding the association between visual attention and predictability of customer shopping behavior.

## III. Proposal

We want to create dynamically changing VR stores, for this we will based on different techniques found on the related work section, our objectives are the following:

- Proactively replacing visible items with hidden ones which might caught more attention from the customers.
- Prioritizing items closely associated with the one currently in focus.
- Dynamically update inventory based on user's visual attention and context search.
- Increase product sales.
- Improve customer experience in a VR store.

In order to achieve this, our model consists mainly on the following sections.

### A. Input

We need a 3D VR store with shelves to place the virtual products, this VR scene will be modeled in Unity and will be possible to interact with it via a HMD device that needs to have eye-tracking capabilities or a simulator. Subsequently we need an inventory of virtual products that matches with the available or valid inventory of a store, at the beginning of the scene only a few products will be displayed and will be changing based on ongoing search context by analyzing the distribution of eye gaze through a FOVE eye-tracking device.

From each product we will take in consideration the following aspects to create a context map model to retain the association rules between items and also be able to predict the related products that an user might be interested in:

- Product ID: Unique identifier for each product.
- Size: Length, width and height.
- Texture: Color and design.
- Brand.
- Price.
- Category: Defined by the administrator.
- Characteristics: Text based description of the product.
- Tags: Defined by the administrator.
- Review Rating: Rating given by customers for the item.
- Height from the floor: This affects how easy the product could be viewed [18].
- Distance from Landmarks: Distance between a product and entrances or the center of the store.

From each user we will take in consideration the following aspects to create an user basic profile to help with the predictability of the related products he or she might be interested in:

- Customer ID: Unique identifier for each customer.
- Age.
- Gender.
- Purchase Amount.
- Location.
- Previous Purchases: List of IDs from previous purchased products.
- Payment Method: Customer's most preferred payment method.
- Frequency of Purchases: Frequency at which the customer makes purchases (e.g., Weekly, Fortnightly, Monthly).

### B. Data Collection

First we need to obtain the data to train our model, for this we have two approaches, one is to conduct an experiment with different user's using the FOVE device in a VR store with multiple products and give them limited virtual money to shop, with this we can obtained the data from the user's attention and also create a dataset for the shoppers behaviour and preferences data. This approach has the inconvenience of being very difficult to implement because, in order to obtain a considerably amount of data we need to run the experiment multiple times and with multiple users with different profiles. This would take a lot of time and budget, for this we plan to use another approach.

The second approach is to use large datasets already created and related to user's shopping context like the ones we can find in **Kaggle website**, clean them, process them, identify what we need and merge it into one that contains all elements needed.

Here we present two data-sets that contains all information needed for our solution:

- **Customer Shopping Trends Dataset**
- **Clickstream Data for Online Shopping**

Combining this two datasets and using data augmentation techniques we can have a really large dataset in order to train our regressor model.

### C. Generate Saliency Maps

We need to define what are we going to define as effective visual attention, we can not use all the data that we get from the eye tracking device since not all of it really caught user's attention, therefore, similar to previous work done by other authors we will only retain if it remains almost stable in a certain location for at least 200 milliseconds. Also we will retrieve all this data once the user finished looking the specific shelve or wall that contains the group of products, followed by this in a background process will generate the saliency map.

With the created map we can also derive into a heat-map of full product, with this the products with more gaze attention generated will enter the context map to obtain related products based on the user's preference and replace this new products with the less attention received ones.

The heat map will use colors from blue to red in order to represent visual time spent, blue will represent zero vision attended to the product and the scale will increase being red the most visually attended products 1.

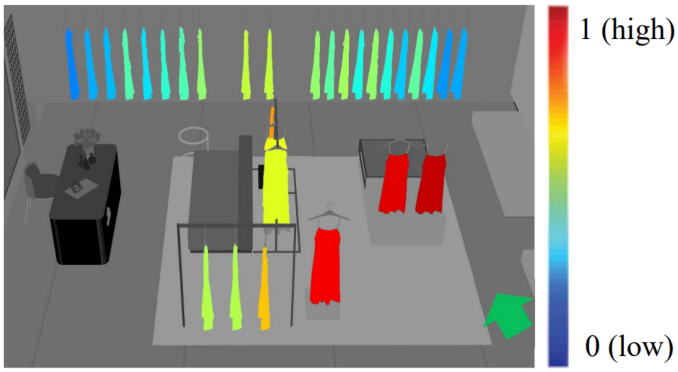


Fig. 1. Visualiztion of heatmap in a VR store scene

### *D. Retrieving data from Saliency Maps*

With the heat map generated in the previous section we create a basic process using a pixel scale detection routine, and we retrieve the top number (will be defined by the administrator) of most visually attended products (known by the scale of colors in the heatmpa), this top products will enter a list of most attractive products to the customer, and the rest will enter the least attractive products list.

### *E. Regressor model for predictability*

With our data obtained we will create a Regression Model to predict consumer future purchases by understanding how customer chooses, buys or discards products. Our solution is based on a Random Forest Classifier model, similarly to what we have found on previous work such as [14] [15] [16]. We are also adding an the association rules to recommend products according to the outcomes, here it is an association rule only based on few attributes of the dataset:

**R1**: **[ProductID: 426569, Brand: Adidas, Color: White, Price: 150, Category: T-shirt]**

**R2**: **[ProductID: 2020426570, Brand: Adidas, Color: Gray, Price: 70, Category: Short]**

**R1 -> R2** indicates that a customer who buys certain white shirt will also buy the matching grey shorts.

To build this association rules we follow an approach similar to [16] where they use of basket analysis to discover the associations between items. The market basket analysis starts with constructing shopping basket data, which is from the purchase combination data set. After our model consumes the results from the heat maps to obtain the IDs of the most visually attended products and identifies an association rules within the data set. Using Apriori algorithm we could find frequent item sets in the data set and analyze them accordingly to establish association rules, then we evaluate the decision data based on these rules, finally, we select the rules with greater confidence (**R2**) to be our recommendations for visually attended product (**R1**).

### *F. Workflow*

Here we present a workflow on how the VR experience will be and the interactions with the previous subsections 2.

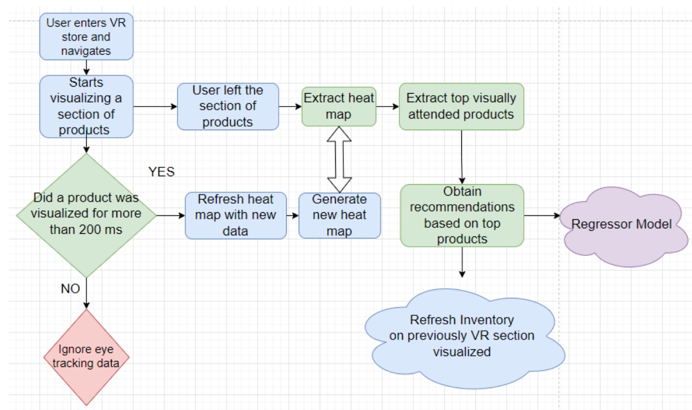


Fig. 2. Worflow of our dynamically changing VR store experience

## IV. Evaluation

We propose an experiment to evaluate if our solution increases sales ratios compared to traditional virtual stores.

The experiment consists in have two groups of user's and let them navigate the VR store with virtual money to spend, they will be able to buy a total of ten products based on their preferences, after the navigation they will be interviewed to know if they were able to find products more related to their preferences in the dynamically changing VR store or in the static one. We also want to know in which store they would buy more often and why. With this data we will be able to know if our approach would increase sales ratio from a dynamically changing VR store compared to an static one, and also see if we improve overall customer satisfaction.

## V. Prototype Demonstration

In the following **link** you can find a demo implementation of our proposal.

## VI. Conclusion

In this work we learned that there is still ground to research on Virtual Reality and how to create better experiences, we found that by applying multiple techniques of computer science we can model a new strategy to do innovative VR projects and that we can implement this classic techniques with a twist for the VR world. With this project we wanted to make VR shopping easier and better for all population since we think there is an added value by experiencing VR stores. For future work we would like to experiment with different approaches to create dynamically changing VR stores and increase the scope where all the aspects of the VR store changes based on user's preferences including the design and textures of the VR itself.

We think that these kind of stores will have a great popularity in the near future and the more acceptance they start to have will guide us to improve the overall experiences.

## References


[1] B. Jin, G. Kim, M. Moore, and L. Rothenberg, "Consumer store experience through virtual reality: its effect on emotional states and perceived store attractiveness," *Fashion and Textiles*, vol. 8, no. 1, p. 19, 2021.

[2] N. Xi and J. Hamari, "Shopping in virtual reality: A literature review and future agenda," *Journal of Business Research*, vol. 134, pp. 37–58, Sept. 2021.

[3] K. K. Evans, T. S. Horowitz, P. Howe, R. Pedersini, E. Reijnen, Y. Pinto, Y. Kuzmova, and J. M. Wolfe, "Visual attention," *WIREs Cognitive Science*, vol. 2, no. 5, pp. 503–514, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.127.

[4] R. Alghofaili, M. S. Solah, H. Huang, Y. Sawahata, M. Pomplun, and L.-F. Yu, "Optimizing Visual Element Placement via Visual Attention Analysis," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Osaka, Japan), pp. 464–473, IEEE, Mar. 2019.

[5] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Jan. 2007. Journal Abbreviation: Eye Tracking Methodology: Theory and Practice Publication Title: Eye Tracking Methodology: Theory and Practice.

[6] S. P. Liversedge, K. B. Paterson, and M. J. Pickering, "Chapter 3 - Eye Movements and Measures of Reading Time," in *Eye Guidance in Reading and Scene Perception* (G. Underwood, ed.), pp. 55–75, Amsterdam: Elsevier Science Ltd, Jan. 1998.

[7] Y. Li, P. Xu, D. Lagun, and V. Navalpakkam, "Towards Measuring and Inferring User Interest from Gaze," pp. 525–533, Apr. 2017.

[8] A. Borji and L. Itti, "State-of-the-Art in Visual Attention Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 185–207, Jan. 2013.

[9] B. David-John, P. Raiturkar, O. Le Meur, and E. Jain, "A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data," pp. 136–139, Dec. 2018.

[10] M. Cerf, J. Harel, W. Einhäuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," vol. 20, Nov. 2007.

[11] C. Guo and L. Zhang, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," *Image Processing, IEEE Transactions on*, vol. 19, pp. 185–198, Feb. 2010.

[12] C. Ozcinar and A. Smolic, "Visual Attention in Omnidirectional Video for Virtual Reality Applications," May 2018.

[13] S. Takahashi, A. Uchita, K. Watanabe, and M. Arikawa, "Gaze-driven placement of items for proactive visual exploration," *Journal of Visualization*, vol. 25, no. 3, pp. 613–633, 2022.

[14] R. I. Karmostaje, "Buyer Prediction Through Machine Learning,"

[15] J. Martínez-Garmendia, "Machine learning for product choice prediction," *Journal of Marketing Analytics*, Apr. 2023.

[16] X. Zhao and P. Keikhosrokiani, "Sales Prediction and Product Recommendation Model Through User Behavior Analytics," *Computers, Materials and Continua*, vol. 70, pp. 3855–3874, Jan. 2022.

[17] T. Germer and M. Schwarz, "Procedural Arrangement of Furniture for Real-Time Walkthroughs," *Comput. Graph. Forum*, vol. 28, pp. 2068–2078, Dec. 2009.

[18] W. Liang, L. Wang, X. Yu, C. Li, R. Alghofaili, Y. Lang, and L.-F. Yu, "Optimizing Product Placement for Virtual Stores," in *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, (Shanghai, China), pp. 336–346, IEEE, Mar. 2023.