Context Aware Recommendations Embedded in Augmented Viewpoint to Retarget Consumers in V-Commerce

Kumar Ayush
Media and Data Science Research
Adobe Inc.
kayush@adobe.com

Abstract

Augmented Reality (AR) has been heralded as the next frontier in retail, but so far, has been mostly used to advertise or market products in a gimmicky way and its true potential in digital marketing remains unexploited. In this work, we leverage richer data coming from AR usage to make re-targeting much more persuasive. Based on the user's purchase viewpoint visual, we identify existing objects/products and recommend products which are stylistically similar to those identified objects and color compatible with the background in the viewpoint. We also embed the recommended products in the viewpoint at the identified objects' location with similar pose and scale. This makes the recommendations much more personalized and relevant which can increase conversions. Evaluation with user studies show that our system is able to make better recommendations than tag-based recommendations, and targeting using the viewpoint is better than that of usual product catalogs.

1. Introduction

Embedding reality in consumers' online shopping experience has been heralded as the "next frontier for retail" and the coming of "v-commerce". V-commerce enables a consumer to overlay a virtual product on the real-world environment to judge its compatibility prior to purchase. Examples include the use of hand-held devices to virtually "try on" furniture/shoes before purchase[1]. AR applications have drawn significant attention in academics [4] and industry[1].

However, these works ignore consumers' preferences necessary to enhance user experience in AR [9]. The proposed approach introduces a robust framework to model visual data generated by AR-based retail apps for targeting. Prior targeting approaches only use information from users' profiles [6], and textual description (content-based model) [11].

A typical AR-based v-commerce app would enable a customer to "tryout" the desired product like chair on a background of her living room. She can either (i) place different chairs on the background, or (ii) move the background around to check the compatibility from different viewing angles. We define viewpoint, to represent the visual at which the consumer judges the compatibility of the virtual product with the surrounding real world environment. The viewpoint holds information previously unavailable from the web-based browsing data, and provides the basis to identify existing products and suggest products similar to the existing ones but with better color compatibility with the background. Also, for enhanced targeting, images of recommended products embedded in viewpoint replacing the identified product/object can be sent.

This paper makes contributions in advancing targeting through AR applications data by:

- Creating persuasive recommendations based on objects/products already present in the user's surroundings represented by the viewpoint.
- Creating personalized catalogues by embedding recommended products in the viewpoint at the location of identified objects with similar pose and scale.

2. Related Work

The deployment of AR in v-commerce enhances consumer experience, as well as provides rich interaction data. The source of AR-based data could be eye-tracking [14], head tracking [15], hand gestures [17] or GPS locations [12]. There has also been significant investment by industry[2] in AR apps. While the IKEA AR catalog app allows customers to have a virtual preview of furniture, Ray-ban's Virtual Mirror enables the consumer to try virtual sun-

[1] www.tinyurl.com/ycvydl89, www.tinyurl.com/yca5krvm

[2] www.ikea.com, www.ray-ban.com

Figure 1. Screenshot frame (left), Camera frame (middle), Identified relevant objects (right). We use only the black armchair for the purpose of recommendations.

Figure 2. Some of the embedded images.

Figure 3. Final Retargeting Images having Embedded Recommendations.

glasses. The rich visual data collected by these apps would help in enhancing consumer experience [7].

In particular, customer viewpoint during an AR app session offers several insights into her preferences. The metric for viewpoint has varied definitions in the literature across different contexts. Vazquez et al. [20] define viewpoint entropy to compute good viewing positions automatically, while [3] shows how to automatically select the most representative viewpoint of a 3D model. An evaluation of the view selection algorithms has been conducted in [5]. However, none of these methods use data from AR-enabled systems for viewpoint selection. [7] is one such work that uses a statistical model to select the viewpoint with the highest likelihood of influencing the consumer's purchase.

Figure 4. Baseline Recommendations.

The customer viewpoint provides a unique advantage to the proposed system over the traditional recommendation systems [2]. The contextual recommendation in [18] exploits users ratings and ontology-based content categorization schemes. Wroblewska et al. [21] rely on images and extract color and texture information to find visually similar items. Our approach can ingest all such data, when available. In addition, the novelty lies in the ability to use viewpoint information to enrich the recommendation.

3. Methodology

The proposed method consists of two stages: (a) Viewpoint Selection (b) Catalog Creation.

3.1. Viewpoint Selection

In Section 1 we defined viewpoint as the visual (image) at which the consumer judges the compatibility of the virtual product (3D model) with the real world surroundings.

There are two challenges that make viewpoint selection difficult: (i) the high volume of images that result from a consumer's session, and (ii) identification of augmented visual(s) from among these sequentially viewed images that the consumer prefers. We use the method employed in [7] to uncover the preferred viewpoint for the consumer. It selects the preferred augmented visual by analyzing the interaction of the consumers and the time stamps at which images (frames) are rendered on the app during a session.

3.2. Catalog Creation

After obtaining the viewpoint, the second step is the catalog creation. For illustration purposes, let the final outcome of our viewpoint selection model be the two (left and middle) images shown in Figure 1. On the left is the AR viewpoint which embeds the virtual table (screenshot image). On the middle is the background viewpoint (the camera image). Unlike [7], we intend to replace an already existing product in the viewpoint with better products. [7] recommends products in place of the one being augmented by the user. The workflow of the recommendation system is as follows:

3.2.1 Object Identification

To create recommendations based on visual information, relevant objects (i.e furniture objects in our case) present in the viewpoint need to be identified (see the right image in Figure 1). We have used Region-based Convolutional Neural Network (R-CNN) [19] which takes as input an image and returns object proposals (bounding boxes) with confidence score and object label. In the running example, we consider the black armchair for replacement with better recommendations.

3.2.2 Alignment of best matching 3D model with an identified 2D object

To create the catalogue, the style and pose of relevant identified objects in the viewpoint is required. For each such object category (for example, chair, sofa, table, etc.), we use an exemplar part based 2D-3D alignment method [1] to find the best matching 3D model from the repository along with their pose and scale in the viewpoint image. For the identified black armchair in the viewpoint, we find the best matching 3D model (in terms of style) along with pose and scale and store it for the subsequent steps.

3.2.3 Style Similarity

To rank recommendations based on their relevance to the user, one criteria that is considered is the "Style Similarity" of a candidate with the identified object. The intuition behind this is that the customer may prefer objects that are structurally similar to the existing model of the identified object. We use the algorithm presented in [13] for this task. We run the algorithm for every candidate model in our repository so as to calculate its style similarity with the identified object/product. For the i_{th} candidate, if d_i is its distance from the identified object model, then we associate a normalized score α_i denoting its style similarity on a scale of 0 - 1 as $\alpha_i = 1/(1 + d_i)$.

3.2.4 Context Aware Removal of Identified Objects and Embedding Recommendations

Next, we know the bounding region for the identified object. We find the precise object mask and remove the object using context aware fill and subsequently embed stylistically similar 3D models present in the repository on the viewpoint at the same location with same pose and scale obtained from the previous step (see Figure 2).

3.2.5 Color Compatibility

Another criteria for ranking the recommendations that is considered is the "Color Compatibility" of a candidate model with the background of the user's purchase viewpoint. To calculate this compatibility measure, we first extract a theme of five colors from the images created in the previous step. This step is done to get a sense of the dominant colors that may attract the attention of the customer. We have used the model presented in [16] which is based on minimizing an objective function that attempts to represent or suggest an image while also being highly rated. For the i_{th} candidate, if t_i is the extracted theme and $r(t_i)$ is the rating of the theme, then we associate a normalized score β_i denoting its color compatibility with the viewpoint on a scale of 0 - 1 as $\beta_i = (r(t_i) - 1)/(5 - 1)$. The user based rating ranges from 1 to 5. Hence, for standardization purposes, the score involves subtracting the rating by minimum possible rating and then dividing it by the difference of maximum possible rating and minimum possible rating.

3.2.6 Overall Score

After devising two normalized scores associated with each possible recommendation as,

- α_i - denoting its style similarity to the best matching model
- β_i - denoting its color compatibility with the user's purchase viewpoint

we formulate an overall score γ_i associated with the i_{th} candidate, assuming it to be a linear combination of the above two scores, i.e., $\gamma_i = w_1 * \alpha_i + w_2 * \beta_i$

Where, w_1 denotes the weight of the style similarity and w_2 denotes the weight of the color compatibility. To compute the weights, we use the Rank-SVM algorithm [10] that employs pair-wise ranking methods. The input to this algorithm is ground truth ranking of the objects having certain scores. It finds out the weights corresponding to each of the scores denoting their importance in ranking. We use a similar approach as [7] and got w_1 = 0.21, and w_2 = 1.65, with cost = 3.5, an accuracy of about 73.19% over validation set and an accuracy of 57.21% on the test set. The magnitude of weights denotes the importance of the features in the ranking (preferences). This also tells us that color compatibility was preferred by the users over style similarity.

3.2.7 Final Recommendations

After computing an overall score for each candidate recommendation, embedded images are ranked based on their overall scores γ_i. We select a fixed number of top ranked images that are to be included in our final catalog (see Figure 3).

4. Evaluation

We evaluate our approach by having humans compare [3] recommendations from our model (Figure 3) with baseline recommendations based on description similarity [8] (see Figure 4). The users were asked to rate each image as Good, Fair or Bad on the following two questions: a) What would be a good recommendation with respect to the identified black armchair in the given living room? (Figure 3 and Figure 4) b) How much engaging the image is containing the recommendation if they were to be sent as email for retargeting purposes? (Figure 3 and Figure 4). We collected 12 responses for each image for each of the above two questions. For the first question (Figure 5 and Figure 6), we found that 87.5% of the times images from our model were rated as either Good or Fair based on how good the recommendations were. Whereas, only 52.083% of the times images from baseline were considered as Good or Fair on recommendations grounds. For the second question (Figure 7), 93.75% of the times images from our model were rated as either Good or Fair based on how engaging the images were for retargeting purposes. Whereas, only 37.5% of the times images from baseline were considered as Good or Fair on how engaging the images were.

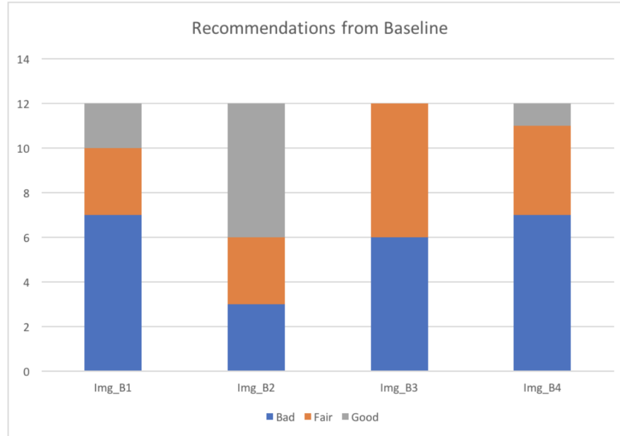

Figure 5. User ratings for recommendations from baseline.

Figure 6. User ratings for recommendations from our model.

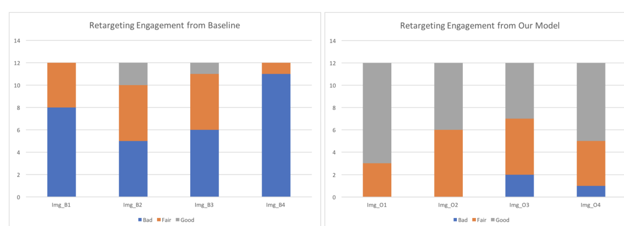

Figure 7. Difference between retargeting engagement of recommendations from baseline (Left) and our model (Right).

5. Conclusion

We create a novel consumer targeting system through modeling the AR-based data. Study tells us that our approach not only produces better recommendations, based on identified objects in the viewpoint, design similarity with those identified objects and color compatibility with the background, but also the recommendations embedded in the viewpoint at the identified objects' locations with similar pose and scale are much more engaging than the usual product images. In future, we plan to deploy this system for comprehensive evaluation, as well as study other context parameters to further enrich the experience.

[3] Internal user study

References

[1] Mathieu Aubry, Daniel Maturana, Alexei A Efros, Bryan C Russell, and Josef Sivic. Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3762–3769, 2014.

[2] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.

[3] Xavier Bonaventura Brugués et al. Perceptual information-theoretic measures for viewpoint selection and object recognition. 2015.

[4] Dameng Dai. Stylized rendering for virtual furniture layout. In *2011 International Conference on Multimedia Technology*, pages 780–782. IEEE, 2011.

[5] Helin Dutagaci, Chun Pan Cheung, and Afzal Godil. A benchmark for best view selection of 3d objects. In *Proceedings of the ACM workshop on 3D object retrieval*, pages 45–50. ACM, 2010.

[6] Lifestyle Finder. Intelligent user profiling using large-scale demographic data. *Artificial Intelligence Magazine. v18 i2*, pages 37–45.

[7] Gaurush Hiranandani, Kumar Ayush, Chinnaobireddy Varsha, Atanu Sinha, Pranav Maneriker, and Sai Varun Reddy Maram. Enhanced personalized targeting using augmented reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 69–74. IEEE, 2017.

[8] Urique Hoffmann, Altigran Soares da Silva, and Moisés G de Carvalho. Finding similar products in e-commerce sites based on attributes. In *AMW*, 2015.

[9] Zhanpeng Huang, Pan Hui, and Christoph Peylo. When augmented reality meets big data. *arXiv preprint arXiv:1407.7223*, 2014.

[10] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.

[11] Przemyslaw Kazienko and Maciej Kiewra. Integration of relational databases and web site content for product and page recommendation. In *Proceedings. International Database Engineering and Applications Symposium, 2004. IDEAS'04.*, pages 111–116. IEEE, 2004.

[12] Bin Liu and Hui Xiong. Point-of-interest recommendation in location based social networks with topic and location awareness. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 396–404. SIAM, 2013.

[13] Zhaoliang Lun, Evangelos Kalogerakis, and Alla Sheffer. Elements of style: learning perceptual shape style similarity. *ACM Transactions on Graphics (TOG)*, 34(4):84, 2015.

[14] Simona Naspetti, Roberto Pierdicca, Serena Mandolesi, Marina Paolanti, Emanuele Frontoni, and Raffaele Zanoli. Automatic analysis of eye-tracking data for augmented reality applications: A prospective outlook. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 217–230. Springer, 2016.

[15] Thomas Nescher and Andreas Kunz. Using head tracking data for robust short term path prediction of human locomotion. In *Transactions on Computational Science XVIII*, pages 172–191. Springer, 2013.

[16] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. Color compatibility from large datasets. *ACM Transactions on Graphics (TOG)*, 30(4):63, 2011.

[17] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*, pages 282–299. Springer, 2013.

[18] Christian Rack, Stefan Arbanowski, and Stephan Steglich. A generic multipurpose recommender system for contextual recommendations. In *Eighth International Symposium on Autonomous Decentralized Systems (ISADS'07)*, pages 445–450. IEEE, 2007.

[19] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[20] Pere-Pau Vázquez, Miquel Feixas, Mateu Sbert, and Wolfgang Heidrich. Automatic view selection using viewpoint entropy and its application to image-based modelling. In *Computer Graphics Forum*, volume 22, pages 689–700. Wiley Online Library, 2003.

[21] Anna Wróblewska and Łukasz Raczkowski. Visual recommendation use case for an online marketplace platform: allegro. pl. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 591–594. ACM, 2016.