Augmented Reality Based Recommendations based on Perceptual Shape Style Compatibility with Objects in the Viewpoint and Color Compatibility with the Background

Kumar Tanmay*
Department of Electrical Engineering
IIT Kharagpur
kr.tanmay147@iitkgp.ac.in

Kumar Ayush*
Department of Computer Science
Stanford University
kayush@stanford.edu

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 via viewpoint image augmentation. Based on the user's purchase viewpoint visual, we identify relevant objects/products present in the viewpoint along with their style such that products with more style compatibility with those surrounding real-world objects can be recommended. We also use color compatibility with the background of the user's purchase viewpoint to select suitable product textures. We embed the recommended products in the viewpoint at the location of the initially browsed product 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 recommendations better 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 [5] and industry[1].

However, these works ignore consumers' preferences necessary to enhance user experience in AR [10]. 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 [7], and textual description (content-based model) [11].

A typical AR-based v-commerce app would enable customer to "tryout" the desired product like a 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 suggest products having better style compatibility with the surrounding real objects and color compatibility with the background. Also, for enhanced targeting, images of recommended products embedded in viewpoint can be sent. This paper makes contributions in advancing targeting through AR applications data by:

- Creating recommendations based on style compatibility with the objects in the Viewpoint and Color compatibility with the background of the user's Purchase Viewpoint.
- Embedding recommended products in the viewpoint at the location of the initially browsed virtual product and 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],

*Equal Contribution
[1]www.tinyurl.com/ycvydl89, www.tinyurl.com/yca5krvm

Figure 1: Screenshot frame (left), Viewpoint camera (middle left), Identified relevant object (middle right), Best matching 3D object of the identified 2D object (right)

Figure 2: Candidate recommendations using Style Compatibility with real world surrounding objects in the viewpoint visual.

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, Rayban's Virtual Mirror enables the consumer to try virtual sunglasses. The rich visual data collected by these apps would help in enhancing consumer experience [8, 2].

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 [4] shows how to automatically select the most representative viewpoint of a 3D model. An evaluation of the view selection algorithms has been conducted in [6]. However, none of these methods use data from AR-enabled systems for viewpoint selection. [8] is one such work that uses a statistical model to select the viewpoint with the highest likelihood of influencing the consumer's purchase.

The customer viewpoint provides a unique advantage to the proposed system over the traditional recommendation systems [3]. 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.

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

3. Data for Recommendations and Targeting Content

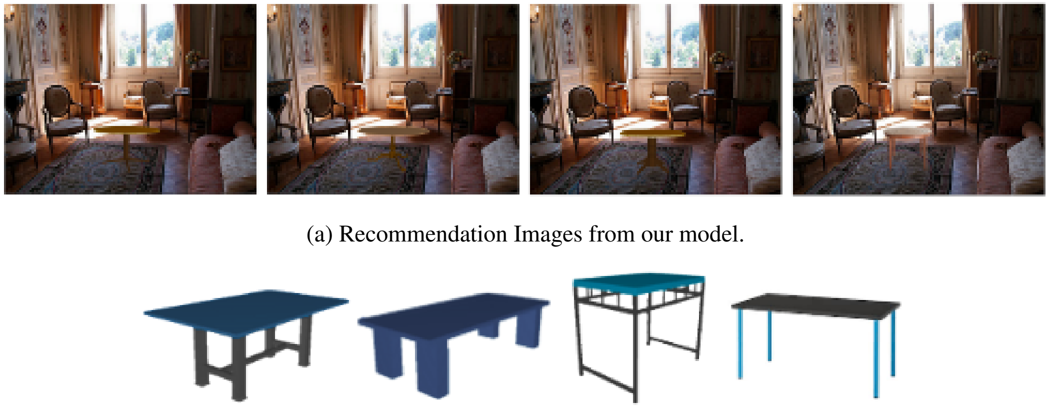
We use an in-house repository of proprietary 3D models of objects which are densely annotated. The dense annotations about real-world dimensions ensures that any tag/description based recommendation system (baseline) has a fair chance to generate good recommendations. For our purpose, we selected a subset of 150 models each from the categories 'armchairs' and 'coffee tables'. The models were selected to form groups based on keyword annotations (design name, color name, etc.) to ensure good recommendation candidates from baseline [9].

4. Methodology

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

4.1. 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 [8] 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.



(a) Recommendation Images from our model.

(b) Recommendation Images from baseline.

Figure 3: Final Catalog with recommendations embedded at the location of the initial augmented product along same pose and scale (a). Tag-based recommendations with respect to the initial browsed product (b).

4.2. 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 left) images shown in Figure 1. On the left is the AR viewpoint which embeds the virtual table (screenshot image). On the middle left is the background viewpoint (the camera image). Unlike [8], we use the styles of already existing products in the viewpoint to recommend products which are stylistically compatible to those existing surrounding real-world objects in the viewpoint. [8] recommends products which are stylistically similar to the one being augmented by the user and does not use the style information of other existing objects in the viewpoint. The workflow of the recommendation system is as follows:

4.2.1 Location and Pose Identification of the Augmented Product being tried by the user

We intend to create a catalogue of images where the recommended products are embedded in the viewpoint image at the location and orientation of the augmented object. Thus, the location and orientation of the augmented object is required in the viewpoint, so that it can be used to embed another object at the same location and in the same orientation. Following [8], we designed our system so that it captures the location and pose of the virtual object in the camera coordinates throughout the consumer's session and then use them for the time point when the viewpoint is selected.

4.2.2 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 middle 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.

4.2.3 Finding the Best matching 3D model of an identified 2D object

For the recommendations, the style of relevant identified objects in the viewpoint is required. For each such identified object and its 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. For an identified object (chair for example) in the viewpoint, we find the best matching 3D model (in terms of style) and store it for the subsequent steps (see rightmost image in Figure 1).

4.2.4 Style Compatibility

To rank recommendations based on their relevance to the user, one criteria that is considered is the "Style Compatibility" of a candidate with the identified objects in the viewpoint. The intuition behind this is that the customer will prefer products that are stylistically compatible with the existing real world objects in the viewpoint. For example, if the user is augmenting a coffee table using the AR app, he/she will prefer a coffee table which is stylistically compatible with a particular sofa, chair, and end table that are currently

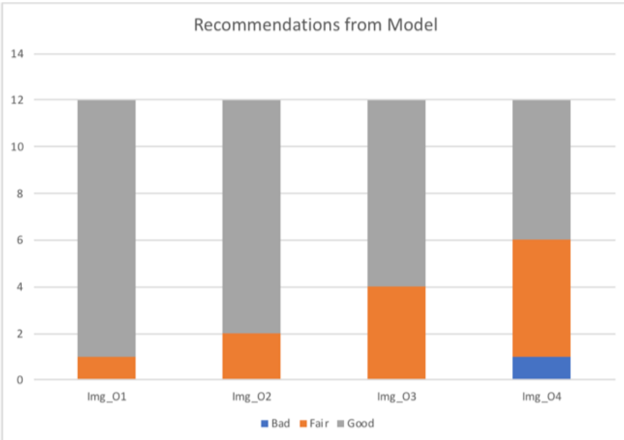


Figure 4: User ratings for recommendations from our model.

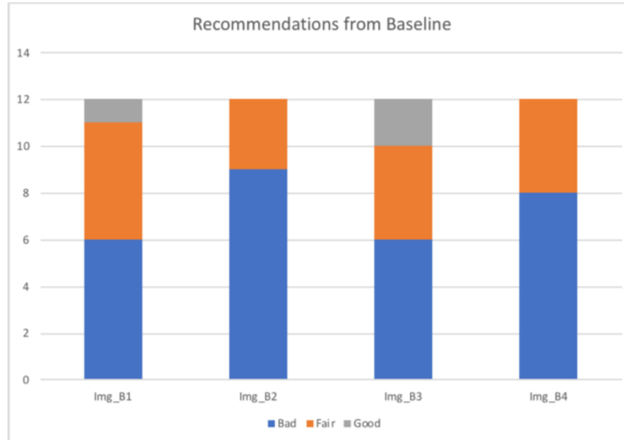


Figure 5: User ratings for recommendations from baseline.

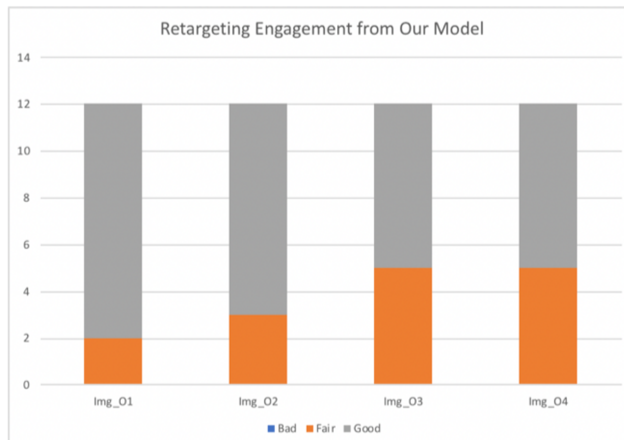


Figure 6: Retargeting engagement of recommendations from our model.

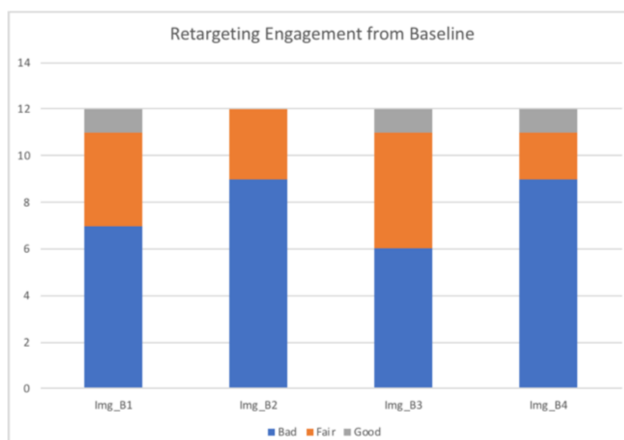


Figure 7: Retargeting engagement of recommendations from baseline.

present in his/her living room/viewpoint. We use the algorithm presented in [13] for this task (see Figure 2).

4.2.5 Embedding Candidates in the Viewpoint

Since we have the camera frame of the viewpoint of the customer along with the location and local orientation of the initial product, we embed all candidate recommendations, obtained from previous step, in their available textures in the extracted frame at the same location with same orientation computed earlier. All candidate recommendations are normalized such that they have the same reference in terms of the rotation, translation and scale. The actual process of embedding could be easily achieved as we have access to relevant APIs of the AR application used by the customer.

4.2.6 Color Compatibility based Texture Selection

Offline shoppers often use color compatibility of the product with the objects in the room. Thus, this criterion is used to select the textures of the products which are color compatible with the background of the user's purchase viewpoint. To measure color compatibility, we first extract a theme of five colors from the created embedded images. 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. Using this score, top textures are chosen to create the final catalogue (see Figure 3a).

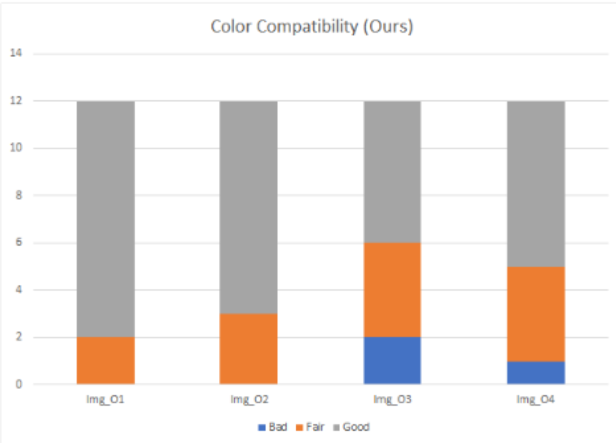


Figure 8: Color Compatibility (with background) of recommendations from our model.

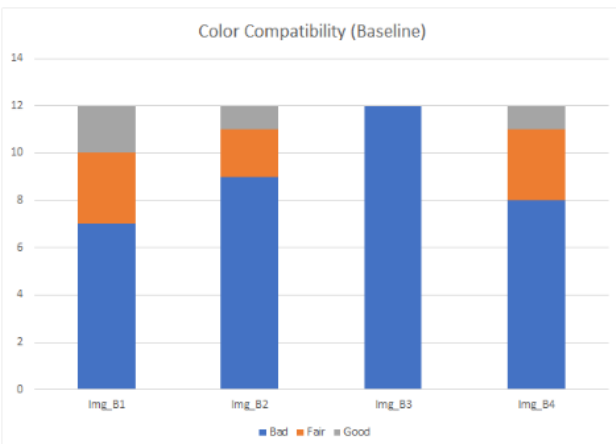


Figure 9: Color Compatibility (with background) of recommendations from baseline.

5. Evaluation

The goal is to have humans compare[3] recommendations from our model (Figure 3a) with baseline recommendations (Figure 3b) based on description similarity [9]. The baseline takes the browsed products as input. It then finds recommendations similar to the input based on attributes such as model, weight, color, etc.

The people in the survey were initially shown the screenshot image (Fig 1 left) and viewpoint camera image (Fig 1 middle left). Then they were shown 8 images (Fig 3) based on the two screenshot and viewpoint camera images. These were the recommendations of products. 4 recommendations {Img_O1, Img_O2, Img_O3, Img_O4} came from our algorithm. Other 4 recommendations {Img_B1, Img_B2, Img_B3, Img_B4} were from the baseline model [9]. The users were then asked to rate each image as Good, Fair or Bad on the following three questions:

- What would be a good coffee table recommendation in the given living room? (Figure 3a and Figure 3b)
- How much engaging the image is containing the recommendation if they were to be sent as email for retargeting purposes? (Figure 3a and Figure 3b).
- How much persuasive the recommendations are based on their color compatibility with the background?

We collected 12 responses for each image for each of the above three questions. For the first question (Figure 4 and Figure 5), we found that 97.91% of the times images from our model were rated as either Good or Fair based on how good the recommendations were. Whereas, only 39.58% of the times images from baseline were considered as Good or Fair on recommendations grounds. Then, we gave weights to the ratings, Good = 2, Fair = 1 and Bad = 0. The weighted average of the ratings for the images from our model on recommendation grounds was 1.71 . This was almost four times the weighted average of the ratings for the images obtained from baseline on recommendations grounds which was only 0.46 . This shows that our images can be considered as in the middle of good and fair ratings, whereas the recommendations from the baseline can be considered as in the middle of bad and fair ratings.

For the second question (Figure 6 and Figure 7), 100.00% 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 35.42% of the times images from baseline were considered as Good or Fair on how engaging the images were. Again, on the engagement for the purpose of retargeting, one can see a clear distinction between the ratings that images generated from our model got and the images generated from baseline got. Then, we gave weights to the ratings, Good = 2, Fair = 1 and Bad = 0. The weighted average of the ratings for the images from our model for retargeting purposes was 1.69 . This was more than four times the weighted average of the ratings for the images obtained from baseline for retargeting purposes which was only 0.42.

Now, on the impact of color compatibility of the recommendation with the background, one can see a clear distinction between the ratings that images generated from our model got (Figure 8) and recommendations from baseline got (Figure 9). 93.75% of the times images from our model were rated as either Good or Fair based on how color compatible the recommendations were with the background. Whereas, only 25% of the times images from baseline (embedded in viewpoint) were considered as Good or Fair based on color compatibility grounds. Then, we gave

[3] Internal user study

weights to the ratings, Good = 2, Fair = 1 and Bad = 0. The weighted average of the ratings for the images from our model on color compatibility grounds was 1.60. This was almost five times the weighted average of the ratings for the images obtained from baseline on color compatibility grounds which was only 0.33.

6. Discussion

While our model performs better than existing methods, it has certain limitations too. Our approach depends on the detection of appropriate viewpoint based on viewpoint selection model in [8]. In some cases, the detected viewpoint might be blurry. Another key bottleneck is the retrieval of matching 3D models of identified 2D objects from the viewpoint image. The retrieved models might not represent the correct style of a particular identified object due to limitations of the algorithm [13], poor resolution images from camera, insufficient repository, etc. Given the limitations, we would like to posit that our work advocates future research towards such personalized recommendations in the domain of augmented reality based virtual commerce.

7. Conclusion

We create a novel consumer targeting system through modeling the AR-based data and creating persuasive visuals for enhanced retargeting. We provide a technology to retarget customers with recommendations which are stylistically compatible with real world objects present in the purchase viewpoint and color compatible with the background. Also, the recommended products are embedded in the purchase viewpoint which leads to further personalization and increases purchase propensity. Evaluation through user studies shows that our recommendations are better than the baseline recommendations, and engagement is improved by embedding the candidates in the viewpoint. In future, we plan to deploy this system for comprehensive evaluation, as well as study other context parameters to further enrich the experience.

References

[1] M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, and J. 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. 3

[2] K. Ayush. Context aware recommendations embedded in augmented viewpoint to retarget consumers in v-commerce. *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, 2019. 2

[3] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013. 2

[4] X. Bonaventura Brugués et al. Perceptual information-theoretic measures for viewpoint selection and object recognition. 2015. 2

[5] D. Dai. Stylized rendering for virtual furniture layout. In *2011 International Conference on Multimedia Technology*, pages 780–782. IEEE, 2011. 1

[6] H. Dutagaci, C. P. Cheung, and A. 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. 2

[7] L. Finder. Intelligent user profiling using large-scale demographic data. *Artificial Intelligence Magazine. v18 i2*, pages 37–45. 1

[8] G. Hiranandani, K. Ayush, C. Varsha, A. Sinha, P. Maneriker, and S. V. R. Maram. [poster] enhanced personalized targeting using augmented reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 69–74. IEEE, 2017. 2, 3, 6

[9] U. Hoffmann, A. S. da Silva, and M. G. de Carvalho. Finding similar products in e-commerce sites based on attributes. In *AMW*, 2015. 2, 5

[10] Z. Huang, P. Hui, and C. Peylo. When augmented reality meets big data. *arXiv preprint arXiv:1407.7223*, 2014. 1

[11] P. Kazienko and M. 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. 1

[12] B. Liu and H. 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. 2

[13] T. Liu, A. Hertzmann, W. Li, and T. Funkhouser. Style compatibility for 3d furniture models. *ACM Transactions on Graphics (TOG)*, 34(4):85, 2015. 4, 6

[14] S. Naspetti, R. Pierdicca, S. Mandolesi, M. Paolanti, E. Frontoni, and R. 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. 1

[15] T. Nescher and A. 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. 2

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

[17] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*, pages 282–299. Springer, 2013. 2

[18] C. Rack, S. Arbanowski, and S. Steglich. A generic multipurpose recommender system for contextual recommendations. In *Eighth International Symposium on Autonomous Decentralized Systems (ISADS'07)*, pages 445–450. IEEE, 2007. 2

[19] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 3

[20] P.-P. Vázquez, M. Feixas, M. Sbert, and W. 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. 2

[21] A. Wróblewska and Ł. 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. 2