Improving Online Customer Shopping Experience with Computer Vision and Machine Learning Methods

Zequn Li, Honglei Li[(✉)], and Ling Shao

Department of Computer and Information Sciences,
Northumbria University, Newcastle upon Tyne, UK
{zequn.li,honglei.li,ling.shao}@northumbria.ac.uk

Abstract. Computer vision and pattern recognition has achieved great developments in last decade, especially the feature categorizing and detection. How to exploit the new techniques in this research area has rarely discussed in the information systems field. This paper aims at exploring the opportunities from the most recent development from computer vision area from the online shopping experience perspective. We discussed the possibility of extracting meaningful information from images and apply this to the online recommendation system to improve online customer shopping experience. Implications to both researchers and practitioners are discussed. The contribution of these papers are twofold, firstly, we have summarized the state-of-the-art of the computer vision development in the online shopping recommendation system, especially in the fashion industry; secondly, we have provided some potential research gaps for on how computer vision method could be used in the information systems field.

Keywords: Online recommendation system · Machine learning · Shopping experience · Image processing · Fashion recommendation

1   Introduction

In recent years, with the development of online shopping, the shopping experience of online customer has been investigated by many researchers [4, 26, 28, 36] from different perspectives. Among these studies an important issue of online shopping experiences lies in the difference between online and offline shopping experience [10, 18]. These researches showed that the socioeconomic variables which traditionally considered being import have changed to be insignificant as before but security aspect tends to be more related. Based on those findings, on-line shopping websites are built to improve the shopping experience from several perspectives including quality control of website [26], interface design for elderly people [27], service quality experience [4] etc. Previous studies have identified that product uncertainty and low retailer visibility will have negative impact on customer satisfaction and thus poor online shopping experiences [33]. Researchers have endeavored to capture more information about products and other features to enhance customers' online shopping experience including utilizing big data, computer vision, and machine learning techniques recently developed.

The online recommendation systems for improved customer online shopping experience have gained popularity because the past transaction data could be used to predict customers purchasing choices [45]. At the same time, there are many successful solutions for online customer recommendation systems [2, 39]. For example, Amazon had increased nearly 30 % of its sales by developing the online recommendation system from customer browsing history. At the same time, the online recommendation system also helps Amazon to control the security and price of the selling item by analyzing the big data provided by customers and products [31, 38]. However, most existing online recommendation systems are developed from readable text [3, 21, 40], leaving many new types of data such as image and multimedia data unused. Multimedia data and image data provides much rich information than readable texts. How to extract meaningful information from multimedia resources like images and apply the extracted information into the online shopping recommendation systems has rarely been considered mainly due to the relevantly new development of computer vision technologies. This study aims at exploring the new techniques from Computer Vision and Machine Learning perspective and proposing a framework of integrating these new techniques with the existing online shopping recommendation systems. The fashion and clothing industry are used in this study as an example to explore such possibility.

We firstly reviewed past research from online shopping experience perspective, mainly on the design of online shopping website or online customer satisfaction, followed by a search for the Computer Vision methods which may be applied to improve online shopping experience. We reviewed most top conferences on Computer Vision such as Computer Vision and Pattern Recognition and ACM Multimedia Conference, especially targeting at the fashion and clothing area. The review results demonstrated that attribute learning method could be used to improve online shopping experience. We illustrated this by demonstrating how fashion item recommendation system could be developed with attribute learning method in computer vision. The implications to both researchers and practitioners are then discussed.

2   Computer Vision Methods and Online Recommendation Systems

The previous research of online shopping behavior [30] shows the dimensions of web site design, reliability, responsiveness, and trust affect overall service quality and customer satisfaction. This paper mainly explore how online shopping experience could be improved from the web site design perspective and explores what new type of technologies from Computer Vision could be used to improve the website design. Meanwhile, we also try to discuss the implications of machine learning and computer vision on information systems theories. We firstly discussed the recent developments in the Computer Vision area and then explore how these new techniques could be applied into the online shopping recommendation system.

2.1 Extract Semantic Attributes from Images

Early studies on online recommendation system rarely consider image as an important factor but only to display the pictures clearly to achieve the optimal product effects [16]. The information in the picture is not fully explored mainly because the image processing techniques haven't been fully developed in early days. Alongside the development of the Image Interactivity technology which enables the creating and manipulation of product images, the potential to exploit more feature from images increase. In the beginning, researchers started focusing on sketching and modelling fashion items [7]. Recently, due to the techniques from machine learning, Computer Vision is witnessed some big breakthroughs. One of the major breakthroughs in Computer Vision is the recognition of image categories [14, 29, 42]. The first improvement comes with feature representation of images, for example, at the feature level, there are kinds of features that could be extracted by different methods including SIFT [24], GIFT [35], Histograms of Oriented Gradient (HOG) [11], Local Binary Pattern (LBP) [1], Maximum Response Filters [43]). Based on these features, a well-trained model could be developed to classify different objects, such as shirts, shoes or hats into categories. The semantic attributes provided by researcher can be used to further assist object classification. Some business solutions had already used this method to preform image mining and achieved satisfactory results [5]. An example of semantic Attribute on Clothes is shown in Table 1.

Table 1. Example of semantic attribute on clothes [5]

Colors	Patterns	Materials	Structures	Looks	Persons	Sleeves	Styles
Beige	Animal print	Cotton	Frilly	Black/white	Child	Long	Nerd
Black	Zebra	Denim	Knitted	Colored	Boy	Short	Outdoor
Blue	Leopard	Fur	Ruffled	Pastel	Girl	None	Preppy
Brown	Argyle	Lace	Wrinkled	Gaudy	Female		Punk
Gray	Check-ered	Leather			Male		Rock
Green	Clotted	Silk					Romantic

However, the problem with this kind of recognition mechanism is that it usually ignores certain type appearance of objects such as the color and texture. In order to solve this problem, some new models were introduced to learn visual attributes [15]. By using this method, human understandable properties could be extracted from images. If we put those properties as labels attached to images, then we can group images by a combination of labels [13, 25]. For example, we can describe a shirt in a specific style with black and white stripes or a white shirt with red round on it and classify clothes with these properties. By using those methods, we could extract some high level semantic features from images such as clothing style, patterns and textures. But these methods only work well with clear and simple image data. As a result, in the realistic online shopping environment, those methods can hardly handle the complex and noisy image resources.

In order to solve this problem, some object detection models have been developed [9, 46]. These models use human pose estimation or simple object detection method to locate the interesting item in an image so that attribute learning method can be applied only to those located item. With this kind of preprocessing method, we could extract semantic attributes from images in a real online shopping environment. There is already some success research in this area. Actually, there is already some success research on this. For example, through collection of a well labelled dataset, Chen et al. [8] extracted complex semantic features from clothing in Fig. 1. Moreover, Liu et al. [32] collected both top and bottom clothes and identified the semantic feature relations between them, which enable them to make further suggestion on item combinations of clothes.

Fig. 1. Extract clothing attribute

As shown above, applying those information collected from Computer vision method could help to improve the design of website and improve not only the description of products but also the shopping experience. However, based on the research of the complexity of website, Park and Kim [36] separated the whole web site into six aspects and find the importance of each part is not equal, and the design of website should not be too complex [17]. So when applying these new technologies into online shopping environment, we need to consider the complexity of new feathers. To apply the huge amount of information provided by Computer Vision methods, certain work is required to be done in information system area to measure the effects of those semantic features. Currently there is no research in the information systems area trying to explore the usage of computer vision methods to improve customer experience. This paper aims at exploring the new perspective and new theories that might arise from the interaction between computer vision and information systems research areas.

2.2   Enrich Recommendation System with Image Features

Analyzing the customers' behavior from their shopping history and using these information to make recommendations to customers so that customers shopping experience could be enhanced has become a trend in most e-commerce websites [10, 36]. Currently, most online shopping websites such as Amazon and eBay make suggestions to their customer by analyzing customers searching or shopping history. This method is successful because related items or products similar to those from their browsed history could be pushed to customers. The limitation is that all the predictions are only based on the item-to-item or user-to-item combinations [31, 37]. The algorithm of these models only considers the relations between item and user or item and item, but ignores the features of the products themselves.

The most salient features extracted from images in e-commerce websites would be used to enhance online recommendation systems and thus shopping experience. Extracted feathers could be those descriptive feathers perceived by human beings such as color and style etc. For example, clothes on Amazon web-site usually contains 5 labels: color, style of sleeve, material and brand, but from the pictures provided by website we can extract more than 10 additional labels such as length, cut, pocket, collar, and material etc. [8, 12]. Moreover, new algorithm could be built based on some public training datasets [5, 23], and well trained model can automatically extract the clothing part and analyzing possible labels from each clothes. These labels could be implemented from human perspectives and some cognitive factors could also be used to extract useful information from clothes pictures. For instance, personality type could be used to classify clothes style based on attributes extracted from images. There is thus a possibility to provide more accurate description of products from higher cognitive and conceptual level so that customers could be provided more enriched products information at higher conceptual and cognitive level.

The overall trend for online fashion recommendation system enables the online shopping systems to be more personalized. There are some successful examples for the fashion recommendation systems through mining the combination of both text and image features. Jagadeesh et al. [22] proposed a fashion recommender by analyzing the color model from street images for item recommendation. Iwata et al. [20] collected text and image data from fashion magazines to build a topic based recommendation system. These two works are item based which only consider the relationship between items and the item-user relationship is not considered here. With the development of social networks, personalized recommendation systems with image features are gaining popularity in recent research. Sigurbjornsson et al. [41] proposed a personalized tag recommendation system based on a Flickr dataset. In this work, they analyzed the frequently used tags of customers to automatically recommend personalized tags for newly added photos. And another research from Yue et al. [47] provided a similar personalize recommendation system by collecting customers' feedbacks. This type of research mostly concentrates on the customer side, and provides recommendations by finding similar customers. Meanwhile, there is also some research considering both user-to-item and item-to-item relationship at the same time. In Hu et al.'s [19] research, they built a model

Fig. 2. Finding tops to match with given bottom and shoes with image features [19]

with each customer's preferred fashion items and then combined these items to make a personalized recommendation for a set of fashion item as shown in Fig. 2.

As shown in Fig. 2, researchers build various recommendation systems through mining the large set of data collected from computer vision methods. However, the current contribution of these new papers is mostly on the new mathematical methods or algorithms that could handle different types of datasets. These works only focus on the recommendation algorithm from the technology perspective. How customers will response to this new type of data hasn't been investigated from the information systems perspective. What type of features shall be extracted? Which features are more salient in improve online customers shopping experience haven't been explored as well.

2.3 Image Analysis with Humans in the Loop

Most Computer Vision problems are solved by machine learning algorithms and there is no need to build a huge image dataset to be learned by that algorithm. Rather researchers need to collect a well labelled fashion dataset for training purpose. The quality of that dataset determined the accuracy of the computer vision model in a certain degree. However, the collection of that dataset is normally expensive and time consuming. Specifically, in fashion and clothing industry, the product and style are changing ever year and fashion companies update their dataset frequently. To solve this issue, the humans in the loop method is proposed [6, 34]. In this method, humans answers are collected for some specifically designed questions, and these questions are formed as human knowledge to enrich the model. Compared with the previous algorithm, the Humans in the loop method use less dataset and get more intelligent results in a dynamic way.

The current progress for humans in the loop methods only have been widely used in animal datasets [6] or unfamiliar classes [44]. There are not any works on fashion items mainly because the feedbacks on fashion items are different among different customer groups, which is not like those structured feedbacks on animals. To improve the humans in the loop methods for the fashion items, more feedbacks from different customer groups could be adopted in the algorithm. The past marketing research findings on customer segmentation could be considered to apply into the humans in the loop methods. The integration of previous marketing theories and information systems theories is expected to contribute to the humans in the loop methods.

3 Conclusion

The purpose of this research aims at exploring the potential to combine the computer vision method with information system method to improve the online shopping experience. We have reviewed and visited a series of computer vision methods and machine learning skills, especially from the fashion area, followed by the current development of online shopping recommendation systems. We found that most online shopping recommendation systems only used the text information from the products and a large amount of information from pictures are not considered in the current online recommendation systems.

We proposed that more fine and enriched information extracted form product pictures with computer vision methods could improve the online shopping experience, and illustrated with the current progress in this area. Although, the potential for the online recommendation system through computer vision methods is very promising there are still many issues to be tackled. We have proposed two important perspectives to be considered to better apply computer vision methods into online recommendation systems. Firstly, what type of semantics features shall be used to build the conceptual models to extract attributes from products pictures? We may have fantastic computer vision techniques but customers may not like any information extracted from product pictures. The conceptual models and even past marketing theories could be used to make the conceptual features more meaningful for computer vision methods. Secondly, with the humans in the loop methods, what type of customer knowledge shall be used to build the algorithm for fashion items?

To apply the computer vision methods into online recommendation system, it's thus essential to gain insights and knowledge from customers' perspective. More research shall focus on testing and investigating customer feedbacks on the current online recommendation systems through computer vision methods. There are also some issues to be solved before applying extracted information from images to the online recommendation system from the technology perspective. Those new algorithms mentioned above all concentrate on the technology side, and most of them only work well with detailed labelled training data. In realistic situation, it might difficult to build the well labelled training data and the images to be analyzed also contain lots of noise data. In this case, the performance of current Computer Vision algorithms should be carefully tested before putting in use.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
2. Andersen, R., Borgs, C., Chayes, J., Feige, U., Flaxman, A., Kalai, A., Mirrokni, V., Tennenholtz, M.: Trust-based recommendation systems: an axiomatic approach. In: Proceedings of the 17th International Conference on World Wide Web, pp. 199–208. ACM (2008)

3. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction for the web. In: IJCAI, vol. 7, pp. 2670–2676 (2007)
4. Bauer, H.H., Falk, T., Hammerschmidt, M.: eTransQual: a transaction process-based approach for capturing service quality in online shopping. J. Bus. Res. **59**(7), 866–875 (2006)
5. Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T., Van Gool, L.: Apparel classification with style. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part IV. LNCS, vol. 7727, pp. 321–335. Springer, Heidelberg (2013)
6. Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., Belongie, S.: Visual recognition with humans in the loop. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 438–451. Springer, Heidelberg (2010)
7. Chen, H., Xu, Z.J., Liu, Z.Q., Zhu, S.C.: Composite templates for cloth modeling and sketching. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 943–950 (2006)
8. Chen, H., Gallagher, A., Girod, B.: Describing clothing by semantic attributes. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 609–623. Springer, Heidelberg (2012)
9. Chen, X., Mottaghi, R., Liu, X., Fidler, S., Urtasun, R., Yuille, A.: Detect what you can: detecting and representing objects using holistic models and body parts. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1971–1978 (2014)
10. Childers, T.L., Carr, C.L., Peck, J., Carson, S.: Hedonic and utilitarian motivations for online retail shopping behavior. J. Retail. **77**(4), 511–535 (2002)
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
12. Di, W., Wah, C., Bhardwaj, A., Piramuthu, R., Sundaresan, N.: Style finder: fine-grained clothing style detection and retrieval. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 8–13. IEEE (2013)
13. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.: Describing objects by their attributes. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1778–1785. IEEE (2009)
14. Fei-Fei, L., Perona, P.: A Bayesian hierarchical model for learning natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 524–531. IEEE (2005)
15. Ferrari, V., Zisserman, A.: Learning visual attributes (2008)
16. Fiore, A.M., Kim, J., Lee, H.H.: Effect of image interactivity technology on consumer responses toward the online retailer. J. Interact. Mark. **19**(3), 38–53 (2005)
17. Guo, Y.M., Poole, M.S.: Antecedents of flow in online shopping: a test of alternative models. Inf. Syst. J. **19**(4), 369–390 (2009)
18. Hernández, B., Jiménez, J., Martín, M.J.: Age, gender and income: do they really moderate online shopping behaviour? Online Inf. Rev. **35**(1), 113–133 (2011)
19. Hu, Y., Yi, X., Davis, L.S.: Collaborative fashion recommendation: a functional tensor factorization approach. In: Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, pp. 129–138. ACM (2015)
20. Iwata, T., Wanatabe, S., Sawada, H.: Fashion coordinates recommender system using photographs from fashion magazines. In: IJCAI Proceedings of International Joint Conference on Artificial Intelligence, vol. 22, p. 2262. Citeseer (2011)
21. Jacobs, P.S.: Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval. Psychology Press, New York (2014)

22. Jagadeesh, V., Piramuthu, R., Bhardwaj, A., Di, W., Sundaresan, N.: Large scale visual recommendations from street fashion images. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1925–1934. ACM (2014)

23. Kalantidis, Y., Kennedy, L., Li, L.J.: Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos, pp. 105–112 (2013)

24. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, p. II-506. IEEE (2004)

25. Kovashka, A., Grauman, K.: Attribute adaptation for personalized image search. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 3432–3439. IEEE (2013)

26. Kuo, H.M., Chen, C.W.: Application of quality function deployment to improve the quality of internet shopping website interface design. Int. J. Innov. Comput. Inf. Control **7**(1), 253–268 (2011)

27. Kuo, H.M., Chen, C.W., Hsu, C.H.: A study of a B2C supporting interface design system for the elderly. Hum. Factors Ergon. Manuf. Serv. Ind. **22**(6), 528–540 (2012)

28. Lai, C.Y., Shih, D.H., Chiang, H.S., Chen, C.C.: The key factors of influence consumer online shopping behavior: using the IQA approach. In: Proceedings of the 8th WSEAS International Conference on E-Activities and Information Security and Privacy, pp. 286–291. World Scientific and Engineering Academy and Society (WSEAS) (2009)

29. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178. IEEE (2006)

30. Lee, G.G., Lin, H.F.: Customer perceptions of e-service quality in online shopping. Int. J. Retail Distrib. Manag. **33**(2), 161–176 (2005)

31. Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **7**(1), 76–80 (2003)

32. Liu, S., Feng, J., Song, Z., Zhang, T., Lu, H., Xu, C., Yan, S.: Hi, magic closet, tell me what to wear! In: Proceedings of the ACM Multimedia, pp. 619–628 (2012)

33. Luo, J., Ba, S., Zhang, H.: The effectiveness of online shopping characteristics and well-designed websites on satisfaction. MIS Q. **36**(4), 1131–1144 (2012)

34. Mensink, T., Verbeek, J., Csurka, G.: Learning structured prediction models for interactive image labeling. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 833–840. IEEE (2011)

35. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vision **42**(3), 145–175 (2001)

36. Park, C.H., Kim, Y.G.: Identifying key factors affecting consumer purchase behavior in an online shopping context. Int. J. Retail Distrib. Manag. **31**(1), 16–29 (2003)

37. Poon, A., Maltzman, R., Taylor, J.: Method and system to recommend further items to a user of a network-based transaction facility upon unsuccessful transacting with respect to an item, US Patent 8,275,673, 25 Sep 2012

38. Rijmenam, M.: How amazon is leveraging big data (2016). https://datafloq.com/read/amazon-leveraging-big-data/517

39. Robillard, M.P., Walker, R.J., Zimmermann, T.: Recommendation systems for software engineering. IEEE Softw. **27**(4), 80–86 (2010)

40. Schmitz, M., Bart, R., Soderland, S., Etzioni, O., et al.: Open language learning for information extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 523–534. Association for Computational Linguistics (2012)

41. Sigurbjörnsson, B., Van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceedings of the 17th International Conference on World Wide Web, pp. 327–336. ACM (2008)
42. Torralba, A., Oliva, A.: Statistics of natural image categories. Netw. Computat. Neural Syst. **14**(3), 391–412 (2003)
43. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. Int. J. Comput. Vision **62**(1–2), 61–81 (2005)
44. Wah, C., Belongie, S.: Attribute-based detection of unfamiliar classes with humans in the loop. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–786 (2013)
45. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)
46. Yamaguchi, K., Kiapour, M.H., Ortiz, L.E., Berg, T.L.: Parsing clothing in fashion photographs. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3570–3577. IEEE (2012)
47. Yue, Y., Wang, C., El-Arini, K., Guestrin, C.: Personalized collaborative clustering. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 75–84. ACM (2014)