https://doi.org/10.2352/ISSN.2470-1173.2017.10.IMAWM-169

Aesthetics of fashion photographs: Effect on user preferences*

Zhi Li**[a]**, Shuheng Lin**[a]**, Yang Cheng**[a]**, Ni Yan**[a]**, Gautam Golwala**[b]**, Sathya Sundaram**[b]**, Jan Allebach**[a]**;
[a]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907, U.S.A;
[b]Poshmark Inc., 101 Redwood Shores Pkwy, 3rd Floor, Redwood City, CA 94065

Abstract

The thriving of online fashion markets has increasingly drawn people's attention. More and more small business owners and individual sellers have joined the traditional professional retail industry, which has led to the blooming of image-based online fashion communities and product photography. Accordingly, we have been dedicated to study how to improve the aesthetic quality of fashion images. In previous work, based on the psychophysical experiments we conducted and the aesthetics evaluation of a given collection of photos, we designed features for aesthetics inference, and introduced a SVM predictor to indicate the image quality. Using this predictor we investigate a large range of fashion photos; and our recent findings show that human aesthetic feedback on fashion images significantly depends on another two high-level factors: the nature of the background in the photo, and how the fashion items are displayed. We believe that fashion photos in which the fashion item is worn by a model, or placed on a mannequin are more aesthetically pleasing than others; and likewise people tend to prefer photos with white background. Furthermore, based on ground truth data that we collected, we perform a statistical analysis to validate these conclusions.

Introduction

When it comes to the growth of e-commerce, the digital fashion industry is the winner of this competition; and it is not going to lose the momentum anytime soon [1]. However, this long-kept momentum also makes online clothing commerce more competitive and fluid. Luxury fashion brands and traditional offline retailers are stepping in this territory; and thanks to this trend, the online customer-to-customer (C2C) fashion market is also thriving. In the C2C commerce mode, users are encouraged to post their own unwanted garments online; and other users can purchase them for a price that is discounted from the cost of the item when it is new. In connection with the blooming of C2C fashion commerce, some problems have emerged. One of the problems we are facing right now is that unlike other professional fashion web retailers, users or photo takers on C2C websites are not professional, and the photos are taken with various low-cost cameras. Therefore, photos on the C2C websites are not guaranteed to have a good aesthetic quality; and the poor aesthetic quality potentially decreases the sales.

Under these circumstances, we want to develop an approach to measure the aesthetic quality of fashion photos. Compared with traditional image quality assessment such as compression artifact analysis, image aesthetic quality analysis would be expected to be more difficult. Some of the foreseeable challenges include the difficulty of mimicking human aesthetics perception, the individual variety of photo aesthetic preferences, and the undercurrent of fashion evolution. However, there have been some revealing works in this area. Jahanian [2] proposed a comprehensive framework for the development of human aesthetics perception based on images. Tong et al. [3] investigated a group of low-level features to determine if images were taken by professionals. Ke et al. [4] introduced a principled method for designing high level features to assess photo quality. Datta et al. [5] extracted certain visual features by intuition and built automated classifiers using a support vector machine and classification trees. Luo et al. [6] proposed a novel method to classify professional and amateur photos by adding subject-related features such as the rule of thirds. This idea was enhanced by Wong et al. [7] by utilizing a visual saliency model. Xue at al. [8, 9] studied the aesthetic properties of photographs and revealed more high-level features.

The project of fashion photo aesthetic assessment was started in our group by Chen et al. [10] in 2014. His successor, Wang et al. [11] introduced an improved framework for the predictor, with more image features and ground truth. He also applied the wrapper method to select the feature subset that yields highest prediction accuracy.

This paper's contribution can be summarized into two points. First, we apply aesthetic quality analysis to an extended range of images, and make some primary observations. Second, two new features, use of garment modeling and use of white background are observed and statistically validated based on the collected image aesthetic quality ratings. To the best of our knowledge, these two new high-level features are not considered as factors for fashion aesthetic quality assessment in any previous fashion photo aesthetics-related paper.

Photo Aesthetics Analysis

In this section, we continue our research based on our SVM-powered aesthetic score predictor with more images from a fashion resell website. We obtain a total of 768 images to create a new image database from 16 different clothing categories. The goal of analyzing the new database is to reveal some common mistakes that amateur photographers make, and to find good practices that can improve the fashion aesthetic quality.

Overall Aesthetic Score Snapshot

We feed the aesthetic quality predictor with the newly built database, and we obtain the histogram of score distribution among all 768 images, which is presented in Fig 1. This figure shows a very pronounced distribution of aesthetic quality scores – they are approximately normally distributed, and centered around 6.5. The standard deviation of this distribution can also be obtained as 0.6, and it follows immediately that usually it is very hard to get an either extremely high (greater than 8.3) or extremely low (less than 4.7) score.

Our goal is to study the images with extremely high and low scores. According to the distribution of the aesthetic scores shown in Fig. 1, all 768 images can be divided into three tiers sorted by two cutoff lines:

1. Tier I: the images with aesthetic scores above 7.14. Because the photos in this tier have the highest scores among all the images, the tier is the best-looking one. The total number of images in this tier is 76, taking 10% of the whole population.
2. Tier II: the images with aesthetic scores between 5.16 to 7.14. This

*Research supported by Poshmark, Inc. Redwood City, CA, 94065

tier takes about 78% of all the images. This tier is regarded as normal-looking, and is of no interest for this paper.

3. Tier III: the images with aesthetic scores below 5.16. The total number of images is 92, which means it contains the worst 12% of images.

Another important aspect is category. The score statistics in terms of different categories are given in Fig. 2. The order of categories is based on the average scores. Jewelry, swim, and accessory are on the top of the ranking. Also, jeans and shorts are the bottom two categories. In fact, most of bottoms other than skirts are ranked below the average. Another interesting observation is that it is easier to get a higher score with smaller items like jewelry, shoes, and bags than it is with larger items, for example, dresses and jeans.

Among all the categories, ”Other” is very special. As shown in Fig. 3, it contains a large variety of items including magazines, candles, music albums, and more. In this case, it is not safe to treat ”Other” as a comparable category because the items do not share any common features.

Worst & Best Tier Studies

Here, we study the worst and the best tiers from the image database mentioned in the previous section.

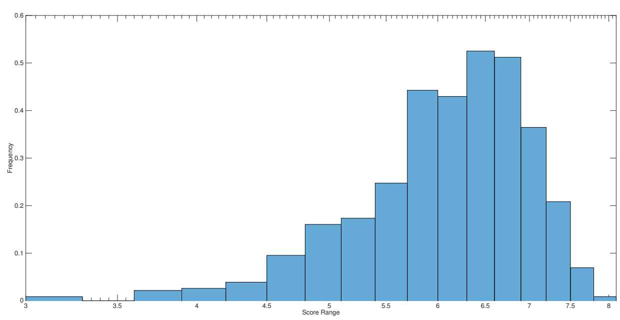

Figure 1. Histogram of overall score distribution. The x-axis is in the scale of log. The distribution is bell-shaped with center around 6.5.

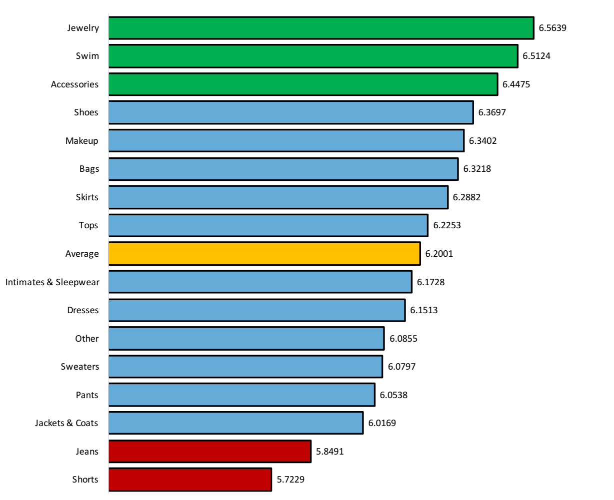

Figure 2. Category comparison. The horizontal axis is the aesthetic score. The orange bar in the middle indicates the overall average score. The top three categories are marked as green. Categories with lowest performances are marked as red.

Table 1 shows the most frequent categories among the worst 12% data. These 4 categories take almost half of all the poor-looking images. Shorts and jeans are both ranked as the worst two categories in both overall ranking and worst 12% ranking, which makes this case very interesting. It will be discussed later. Even though 14 or 13 occurrences in the worst looking tier of photos, from these two categories may seem reasonable, it should be noted that only 48 photos were downloaded from each category. Therefore, these numbers represent more than 25% of the photos in each of these two categories. On the other hand, there are also 9 items from the ”Other” category in this worst 12% group, but as stated in the previous section, it is not listed in the Table 1; and our discussion will not cover this category.

Table 1: Category distribution of images in worst tier

Occurrence	Category(s)
14	Shorts
13	Jeans
8	Dresses
7	Intimates & Sleepwear

Table 2 indicates the most frequent categories in the best tier. These top-listed frequent categories have very similar occurrences. The category of dresses occurs in the both best tier and worst tier. So it has a large variation. It is safe to say from this table smaller items like makeup, jewelry, and shoes, it is usually easier to get a higher score than it is with the larger items like jeans and pants.

Table 2: Category distribution of images in best tier

Occurrence	Category(s)
8	Dresses
7	Makeup, jewelry, swim
6	Tops, shoes, accessories, skirts

After seeing the statistics, we analyze some pictures from these two tiers. Figure 4 shows a group of images in the worst tier. Here, we have

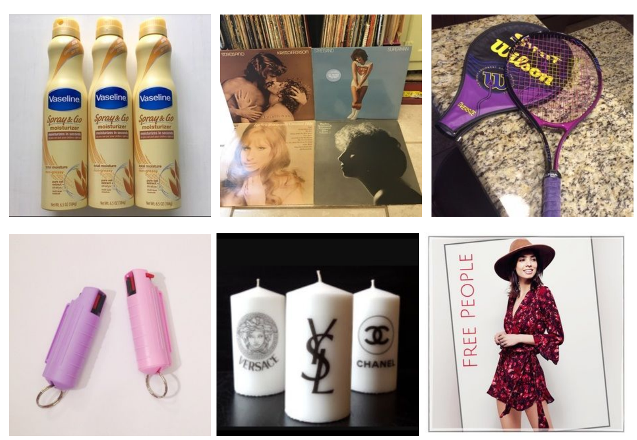

Figure 3. Examples of photos in ”Other” category. Items in this category do not share any common features. The predicted aesthetic scores for these images are: 5.87, 5.15, 7.09, 6.54, 4.96, 5.16 (left to right, then top to bottom).

shorts, jeans, pants, and dresses. Firstly, from this set of images, we find that users tend to lay items against a platform horizontally (for example, lay on a bed or floor) or vertically (for example, hang on a wall). This usually does not lead to a good result – in this case, a good aesthetic quality/score. One of the possible explanations is that laying items on a surface results to the deformation of the clothes due to the soft and flexible nature of the fabric. That means that images of clothing items that are designed for usage in 3D space, will lose information when projected to a 2D surface. Second, most of the garments in this tier usually only have one color; and these selected colors are very dark and saturated. This is not very aesthetically pleasing, either. In clothes with dark pure color it is difficult to present high frequency details, and easy to be distracted by the background. This might be an natural disadvantage of those clothes with these colors. Another observation we can make here is that most of the photos in the worst tier have a dark, or very busy background; and this type of background doesn't help to improve the scores.

Figure 4. *Examples of photos in worst tier. Items are tend to have pure color, and are exhibited against a flat surface. The predicted aesthetic scores for these images are: 4.83, 3.64, 4.97, 4.33, 3.81, 2.65 (left to right, then top to bottom).*

Compared with the worst tiers, Fig. 5 shows some examples from the best looking tier. As we can see, most of the items have a bright and clean background. The clothes are properly modeled or displayed using a mannequin, while bags and shoes do not need to be placed on a mannequin or modeled because they keep their shape more easily than clothes.

As stated above, a white background is a good aid to improve aesthetic quality. In fact, it is also a common practice in the product photography industry. According to Pixelz, an image post-editing company [12], a white background gives images better consistency and more accurate color representation. Furthermore, aesthetic quality is not the only the benefit of white background, several other advantages are listed below:

- White background eliminates all possible distractions from the garment, so it can help the viewers concentrate on the garment.
- White background can reduce the file size during image compression.
- White background is also widely used in modern webpage design, so garment photos can blend better with the website.
- White background can reinforce the color accuracy of the garments or other products.

However, the definition of white background needs to be well specified before any generalization. According to the Retinex theory [13], the human perception of color is subject to the surrounding environment, and this theory also holds when talking about amateur product photography. The color of white always refers to the (#FFFFFF) white for studio photography where professional equipment such as an infinity white cove [14] is used. But for amateur photo shooting, users generally are not able to provide such an equipment, and various kinds of backgrounds are deployed. Therefore, we generalize the term "white background" into a broader spectrum when dealing with the amateur photography: if the background of an image is not cluttered, has a pure neutral color, and approximates the functionality of a studio white background, then we call it a white background.

In Fig. 6, some examples of photos with white background are given. All three images have different color tones (especially the tone of the first picture is heavily modified). But the backgrounds have less high frequency details and distraction, so we consider these photos to have a white background.

For comparison, we also have a group of photos in Fig. 7 to show what photos with non-white background look like. In these three photos, the actual garments to be presented are over shadowed by the background. The handbag in the first photo and the jacket in the second photo are displayed in front of a very busy background; and the lighting is not highlighting the garments, either. The background in the last photo is not as busy as the left two, but the color of the background is close to the garment itself, so it is very hard to distinguish the item from the whole image.

Another very important aspect is the usage of a model or mannequin. As discussed before, without proper support, soft clothing items will lose their shape and original design. Therefore, having someone try it on is an easy way to bring the garment to life. In Fig. 8, some other examples of

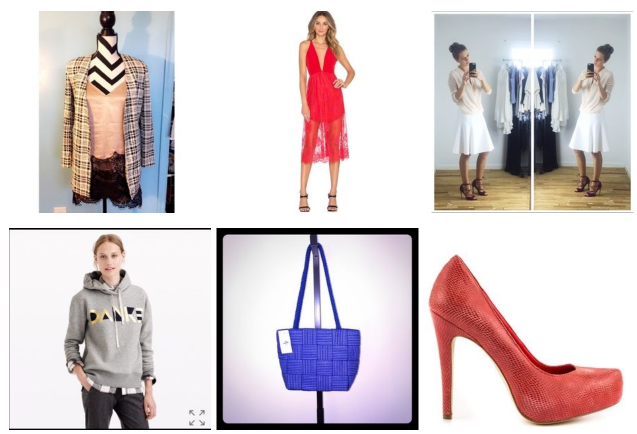

Figure 5. *Examples of photos in best tier. Most items are properly modeled or supported by a mannequin, and a white background is widely used. The predicted aesthetic scores for these images are: 7.84, 7.41, 7.37, 7.24, 7.92, 7.64 (left to right, then top to bottom).*

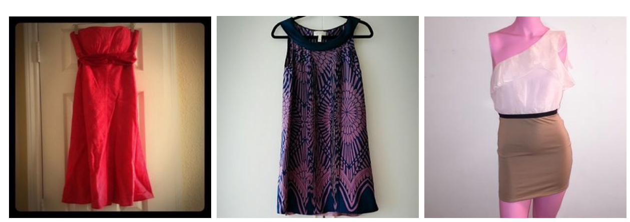

Figure 6. *Examples of photos with white background. The predicted aesthetic scores for these images are: 5.27, 5.34, 6.80 (left to right).*

modeled garments are given. This type of aids effectively helps to convey the impression of the garment, and it highlights the composition of the photo, too.

To summarize the discussion above, the observations can be described as follow:

- white or neutral background might be able to improve the aesthetic quality of fashion garment photos
- models and mannequins are a great aid to showcase the garment, gaining a better aesthetic quality.

Real World Rating Testing

In this section, further testing is conducted to verify the previous observations:

1. Modeled garments have better aesthetic quality than unmodeled garments.
2. Fashion photos with white background have better aesthetic quality than those of other background.

In order to verify the observations above, two problems need to be addressed. First, as mentioned above, both observations are drawn based on the results of SVM-powered aesthetic quality predictor, which simulates the human perception of aesthetics. But in order to validate the statements above, real world ratings produced by human subjects are needed. Secondly, proper statistical testing methodologies should be implemented.

Ground Truth Collection

Instead of the set of images used in the previous section, a new independent dataset should be used. Here, we utilize a previously collected image dataset that contains 734 fashion photos. Each of the 734 photos is rated by multiple female viewers; and we calculate the average of all the individual ratings as the final rating [11].

Addition to the rating collection, we invite a female student to label all images as modeled/unmodeled and white-background/other-background; and the results are shown in the Tables 3 and 4.

As we can see, the data are not balanced between modeled and unmodeled, which is the same as the white background/other background.

Figure 7. *Examples of photos with non-white background. The predicted aesthetic scores for these images are: 6.31, 6.40, 4.76 (left to right).*

Figure 8. *Examples of photos w/ modeled garment. The predicted aesthetic scores for these images are: 7.10, 7.02, 6.82 (left to right).*

Table 3: Statistics of modeled and unmodeled sample sets based on ground truth ratings

Sets	Modeled	Unmodeled
Set size	214	520
Sample mean	6.5432	5.4116
Sample standard deviation	1.8074	1.5756

Table 4: Statistics of white background and other background sample sets based on ground truth ratings

Sets	White	Other
Set size	604	130
Sample mean	5.5571	6.5672
Sample standard deviation	1.2931	1.3074

Besides, both the modeled and white background photos have approximately a 1 unit advantage in mean aesthetic score over the unmodeled and other background photos, respectively.

Hypotheses Testing

First, consider the difference between modeled/unmodeled photos. For Fisher's hypothesis test, two hypotheses should be given before processing the data [15]. Therefore, both the null hypothesis H_0 and the alternate hypothesis H_1 are given as below:

- H_0: The average rating of modeled photos is same as that of unmodeled photos
- H_1: The average rating of modeled photos is higher than that of unmodeled photos

Note that here the alternative hypothesis is defined as one-sided. Let μ_0 be the average rating of modeled photos, and μ_1 be the average rating of unmodeled ones. Then both hypotheses can be rewritten as

- H_0: $\mu_0 = \mu_1$
- H_1: $\mu_0 > \mu_1$

With the hypotheses being declared, statistics of samples should be obtained. The empirical standard error, which we abbreviate SE, can be written as

$$\mathrm{SE} = \sqrt{\left(\frac{1}{n_0} + \frac{1}{n_1}\right) \frac{(n_0-1)s_0^2 + (n_1-1)s_1^2}{n_0+n_1-2}} \tag{1}$$

where n_i is the size of i-th sample set; s_i^2 is the sample variance of sample set i

$$s_i^2 = \frac{1}{n_i} \sum_{k=1}^{n_i} (r_k^{(i)} - \bar{r}_i)^2 \tag{2}$$

and $\bar{r}_i$ represents the corresponding sample mean of sample set i

$$\bar{r}_i = \frac{1}{n_i} \sum_{m=1}^{n_i} r_m^{(i)} \tag{3}$$

Note that $i = 0$ or 1, where 0 indicates that the modeled group, and vice versa. The numbers are shown in the Table 3 for these two sets of interests.

Then, we can obtain the SE value of 0.1416. After this, the T-score, which is also called *Welsh's t*, can be calculated by

$$t_d = \frac{\bar{r_1} - \bar{r_2}}{\text{SE}} \tag{4}$$

The t-value in this case is 7.9654, and the corresponding p-value is 2.3352×10^{-14}. Therefore, for confidence level $\alpha = 0.01$, we reject the null hypothesis in favor of the alternative.

Similarly, the two sample t test is implemented to examine the relationship between neutral background and aesthetic quality, which, as before, is represented by human subjects' ratings. The results are given in Table 4.

The results of t-test are: the t-value is 6.2334, and the p-value is 3.0114×10^{-9}. Therefore, we reject the null hypothesis in favor of alternative with confidence level $\alpha = 0.01$.

Conclusion

In this paper, our previously developed SVM-powered photo aesthetic quality predictor is applied to a set of amateur fashion photos, and the results of follow-up studies are presented. We obtain 768 photos from an online fashion resale website that cover 16 women's garment categories, and run them through the predictor. Based on the scores produced by the predictor, two observations are drawn: (1) Aesthetic quality scores are approximately normally distributed; and it is very hard to get either extremely high or low aesthetic scores; (2) With fashion photos of larger garment pieces, such as jeans, jackets and dresses, it is more difficult to achieve high aesthetic quality than with those of smaller pieces like jewelry and accessories.

By studying some cases where the predicted scores are in the lowest 12%, we also find out that photos of clothes with pure dark color are prone to a lower score. Another common feature among these cases is related to the photographers' techniques and preparation – amateur photographers don't usually use a mannequin or live model; and they place their items in front of a very busy and not light background. Therefore, we make a hypothesis that garment modeling and a light and neutral background improve image aesthetic quality. Furthermore, we conduct a T-test with another group of 734 photos with corresponding human subjects' ratings, which provides further support of the above hypothesis.

Acknowledgments

We thank Poshmark Inc. for their continued support on our research project, and following people for their contributions and suggestions regarding this work: Apoorv Ankulam, Franklin Shiao, Roro Okpu, Nikhila Beesetti, and Kartikeya Mishra.

References

[1] J. Love, "Behind the online apparel boom," 2016. [Online]. Available: https://www.internetretailer.com/2016/07/01/behind-online-apparel-boom

[2] A. Jahanian, *Quantifying aesthetics of visual design applied to automatic design*. Springer International Publishing, 2014.

[3] H. Tong, M. Li, H.-J. Zhang, J. He, and C. Zhang, "Classification of digital photos taken by photographers or home users," in *Proceedings of the 5th Pacific Rim Conference on Advances in Multimedia Information Processing - Volume Part I*, ser. PCM'04. Berlin, Heidelberg: Springer-Verlag, 2004, pp. 198–205.

[4] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, June 2006, pp. 419–426.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proceedings of the 9th European Conference on Computer Vision - Volume Part III*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 288–301.

[6] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Proceedings of the 10th European Conference on Computer Vision: Part III*, ser. ECCV '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 386–399.

[7] L.-K. Wong and K.-L. Low, "Saliency-enhanced image aesthetics class prediction," in *Proceddings of the 2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 997–1000.

[8] S.-F. Xue, Q. Lin, D. R. Tretter, S. Lee, Z. Pizlo, and J. Allebach, "Investigation of the role of aesthetics in differentiating between photographs taken by amateur and professional photographers," in *Proc. SPIE*, vol. 8302, 2012, pp. 83 020D–83 020D–5. [Online]. Available: http://dx.doi.org/10.1117/12.914686

[9] S. F. Xue, H. Tang, D. Tretter, Q. Lin, and J. Allebach, "Feature design for aesthetic inference on photos with faces," in *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP)*, Sept. 2013, pp. 2689–2693.

[10] M. Chen and J. Allebach, "Aesthetic quality inference for online fashion shopping," in *Proc. SPIE*, vol. 9027, 2014, pp. 902 703–902 703–7. [Online]. Available: http://dx.doi.org/10.1117/12.2045269

[11] J. Wang and J. Allebach, "Automatic assessment of online fashion shopping photo aesthetic quality," in *Proceedings of the 22nd IEEE International Conference on Image Processing (ICIP)*, Sept. 2015, pp. 2915–2919.

[12] Dave Kensell, "The Pixelz Product Image Report: What 7 Million Edited Images Can Teach You [Infographic]," 2016. [Online]. Available: https://www.pixelz.com/blog/what-7-million-edited-images-can-teach-you-infographic/

[13] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 1–11, Jan 1971. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=josa-61-1-1

[14] Shootfactory, "What is an infinity curve or cove in photography," 2015. [Online]. Available: http://www.shootfactory.co.uk/what-is-an-infinity-curve-or-cove-in-photography/

[15] T. Snijders, *Hypothesis Testing: Methodology and Limitations.* Elsevier-Pergamon, 2002, pp. 7121 – 7127.

Author Biography

Zhi Li is a third year doctoral student and research assistant in School of Electrical and Computer Engineering at Purdue University, West Lafayette. He is working with Prof. Jan Allebach and Poshmark primarily on fashion photography aesthetics, studying how photography techniques change human perspective towards image aesthetic quality. Besides the image aesthetic research, other fashion-related topics also interest him; a garment color extraction project is ongoing; and he is also developing a fashion photo categorization method using deep neural network.