# Climaxing VR Character with Scene-Aware Aesthetic Dress Synthesis

Sifan Hou*
Beijing Institute of Technology

Yujia Wang†
Beijing Institute of Technology

Wei Liang‡
Beijing Institute of Technology

Bing Ning§
Beijing Institute of Fashion Technology

Figure 1: Our approach automatically synthesizes aesthetic dresses for the virtual character in accordance with different scenes. The synthesized clothes are in harmony with both scene semantics (e.g., season, occasion, and category) and character's appearance attributes (gender, age, figure, and height).

## Abstract

Like real humans, virtual characters also need to dress up according to different application scenarios so that the virtual character appears professionally, harmoniously, and naturally. However, manual selection is tedious, and the appearances of virtual characters usually lack variety. In this paper, we propose a new problem of synthesizing appropriate dress for a virtual character based on the scenario analysis where he/she shows up. We come up with a pipeline to tackle the scenario-aware dress synthesis problem. Firstly, given a scene, our approach predicts a dress code from the extracted high-level information in the scene, consisting of season, occasion, and scene category. Then our approach tunes the dress details to fit the aesthetic criteria and the virtual character's attributes. An optimization of a cost function implements the tuning process. We carried out experiments to validate the efficacy of the proposed approach. The perceptual study results show the good performance of our approach.

**Index Terms:** Digital Fashion; Visualization Design and Evaluation Methods; Fashion Outfit Generation

## 1 Introduction

Virtual characters in Virtual Reality and Mixed Reality not only enrich users' experience, but also assist people to complete various tasks in real world. A virtual character could be an intelligent personal assistant to help users with administrative tasks; could be an avatar to represent a user in a virtual world; could be a chatbot to present itself. In order to optimize the users' feelings and experience, improving the authenticity of virtual characters has gradually become an important topic of mixed reality, a field about how to introduce the virtual world into the real world. The development of Virtual Reality and Mixed Reality facilitates the virtual characters to appear in kinds of scenarios, playing the role of virtual assistants, virtual friends and so on. Taking a virtual assistant as an example, he/she often attends formal occasions, such as meetings and ceremonies. The broad applications of virtual human technology bring a new problem: when a virtual character appears, what should he/she wear? For a person, dressing for any occasion is about paying respect to the occasion and other people, so is a virtual human.

To ensure a virtual character to appear with appropriate outfits in different scenes, a Mixed Reality system for outfits synthesis based on scene understanding is highly desired. There are two ways to dress up a virtual character in current applications: dressing the virtual character in a default suit or providing skin for the user to select manually. Obviously, always dressing a virtual human in the same suit on any occasion is not proper, which may affect human experiences. For example, for a virtual personal assistant, when the user needs some professional suggestion in a business, it is proper for the virtual personal assistant to wear a business suit. Differently, in a daily scenario, it is more common to wear a casual dress. Manual selection is another option, but it is tedious and prohibitive on a large scale if the selection occurs every day.

In this paper, we propose a new task of synthesizing an appropriate dress for a virtual character automatically. The problem of dressing up a virtual human is similar to the process of selecting dresses for a real person. A dress should meet the requirements of both functionality and aesthetic rules. People generally use dress code to discriminate the functionality of different dresses, while the dress looking is also highly related to the detailed parts. Thus, we consider two important factors in the dress decision for a virtual character: dress code selection (dress category) and aesthetic dress optimization (dress appearance).

Dress code is usually created out of social perceptions and norms, and varies based on tasks, circumstances, and occasions. Some previous works [12, 24, 38] have demonstrated the correlation between the dress code and some high-level information conveyed by scenes, e.g., season, occasion, scene category. It is intuitive to predict the dress code from the scene where the virtual character will show. To tackle the prediction problem, we model the relation between scenes and dress code by a fully connected neural network. The scene is represented by the extracted high-level information; and the dress code is represented by a set of dress categories. Given

*e-mail: 3120191002@bit.edu.cn
†e-mail: wangyujia@bit.edu.cn
‡e-mail: Corresponding author, liangwei@bit.edu.cn
§e-mail: Corresponding author, ningbing@bift.edu.cn

a scene, the learned neural network can predict a dress code to be compatible with the scene automatically. According to the dress code, our approach subsequently chooses suitable outfits for virtual characters.

Besides the dress code, a dress looking is comprised of many detailed parts, e.g., collar shape, sleeve length. Those dress details reflect the aesthetics of the dress and highly depend on a virtual character's attributes, e.g., gender, height. We formulate the details synthesis as an optimization problem, taking the virtual character's attributes into account. An aesthetic evaluation model is learned and is applied to guide the optimization process to search for an optimal dress details configuration. Our framework is flexible and scalable, allowing more semantic information, constrains of virtual characters, dress attributes and user's interaction.

Our approach synthesizes a proper dress for a virtual character according to the scenario by two steps, i.e. dress code prediction and aesthetic dress synthesis. Given a scene image, the corresponding high-level information is extracted from the scene, including season, occasion, and scene category. Then our approach tunes the dress details according to the virtual character's attributes, so that the synthesized dress satisfies the aesthetic rules, and matches the virtual character's appearance. We demonstrate some synthesis examples in Figure 1. In each group, the left column is the input scene, from which our approach extracts the scene semantics to guide the synthesis. The right column is the dressed virtual characters, who wear the synthesized dresses.

The main contributions of our paper are as follows:

- Propose a new problem of dress synthesis according to scene information and character features.
- Devise a computational framework based on *Scene-Relevant Dress Dataset*, which explicitly leverages the relationship between scene semantics, character features and dress attributes.
- Experiment the proposed approach on different scenes and virtual characters; validate the effectiveness through perceptual studies.

## 2 Related work

**Fashion Design** Dress is a common topic in people's daily life, and brings dress design gradually becoming an important domain in computer vision and multimedia communities. Some researches on dress understanding, such as clothing parsing [32, 34], visual feature extraction [16], style understanding [17, 19] and fashion compatibility prediction [10], provide a good foundation for dress design.

Fashion design aims to design styles and details for every clothing item according to some specific needs. There exist few works on designing an outfit or a clothing item relatively, but there are already some studies on the partial design of clothes. A particular local feature of a clothing item may inspire people' love for it. For example, a bubble collar will make a rigid shirt look cute. This has made attribute editing an emerging field of fashion design. [21] introduced a novel task into the fashion domain named attribute-aware fashion-editing, which edits certain attribute(s) of the image of a fashion item and preserve other details as intact as possible. [4] innovatively proposed to graft a certain attribute of a clothing item to another clothing item, and generated a photorealistic image that combines the texture from reference image and the new attribute from another image. Fashion design is the process of transforming requirements and inspiration into clothing entities. Considering that needs and inspiration can be communicated through text, [2] proposed an approach named eAttnGAN learned from the FashionGen dataset [22], which can generate images based on text descriptions.

All these ideas and technologies open a new door of possibilities for user-driven fashion design, and are potentially beneficial to virtual try-on, clothing recommendation, visual search, etc.

**Dress for Occasions.** Every day, we have to face such a question: what should I wear today? This is often influenced by the circumstances of the day: what is the weather like today, where I am going, and what kind of occasion it should be? [12] revealed that outfit choices are driven by a variety of factors, including scene factors, e.g., seasonal, occasional and so on.

There has been much research dedicated to enabling users to easily organize and optimize their daily clothing selection based on scene needs. Combining multiple scenario-related options such as the user's planned activities and the weather with other required information, [26] helps the user coordinatie what to wear. [15] proposed a latent Support Vector Machine model for occasion-oriented clothing recommendation, that is, given a user-input occasion, recommending the most appropriate clothing, or recommending items to match with the reference clothing. Through the data observation of the photos from several popular travel websites, [38] found that people's choice of clothing items and their color combinations have strong correlations with the weather, the season and also the main type of scenic spots at the destination, so they learned the relationship between clothing and locations from social photos and applied it for location-oriented clothing recommendation. [24] analyzed the fashion of clothing with a conditional random field model that considers the setting of photos such as the scenery behind the user.

It is obvious that scene information has a great influence on people's dress. The season determines the thickness of our clothes. For example, in summer, we just need a t-shirt and a pair of cool pants, while in spring and autumn, we should add a coat, not to mention that we need to wear a sweater under the down jacket in winter. Besides, scenes and occasions have more detailed requirements for dress codes. The athletes on the football field should wear sports suits, and the host of the ceremony should wear gorgeous dresses. In daily life, people often wear casual and comfortable clothes.

**Dress Synthesis** The rapid development of computer vision and computer graphics provides people with the possibility of automatically generating clothing. Strong technologies like clothing modelling and simulation have helped people realize the real-time visualization of clothing, and are still in the process of improvement.

One of the research directions of clothing generation is matching, which has led to much research on clothing recommendation. By analyzing fashion attributes and learning fashion compatibility from massive images, some algorithms have a deep understanding of fashion; thus they can recommend clothes with great fashion compatibility for users [14]. Considering the personalized requirements of people today, some algorithms add people's fashion preferences as attributes into the recommendation system, so they can make recommendations that meet the user's taste when people buy clothes or choose what to wear every day [10, 26, 36].

Meanwhile, some researchers are committed to improving the authenticity of clothing simulation, such as enabling the simultaneous animation of thousands of detailed garments in real-time [1, 31], endowing the individuals with realism and variety [5], animating realistic clothing on synthetic bodies of any shape and pose without manual intervention [8], defining models for knits in terms of the motion of yarns, rather than the motion of a sheet [13]. [3] describes a cloth simulation system that can stably take large time steps. [7] introduces a hybrid method for realtime cloth animation. [30] proposes a method which supports the production of high quality textured garment designs by optimizing the continuity and symmetry properties of the texture across seams. [27] presents a novel interactive tool for garment design that enables interactive bidirectional editing between 2D patterns and 3D high-fidelity simulated draped forms.

Previous studies on clothing synthesis mainly focused on selecting and recommending clothing to real human users from existing clothes based on aesthetic criteria [15], wearing preferences [11], tourist attraction attributes [38]. These works lack understanding of

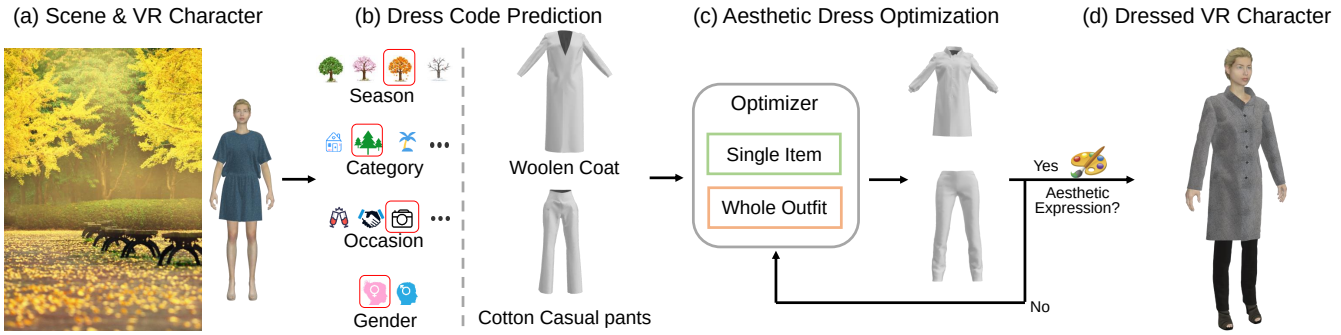


Figure 2: Overview of our approach. Given a scene and a virtual character (a), our approach synthesizes an aesthetic dress for the virtual character through two stages, i.e. (b) dress code prediction and (c) aesthetic dress optimization. The dress code prediction determines the dress type. Then the aesthetic dress optimization adjust the dress details. The synthesis result is shown in (d).

scene semantic and consideration of virtual characters' attributes, and are limited to the category and quantity of existing clothing items [19, 26, 36]. Differently, based on existing works of dressing for occasions, which is closely related to dress code prediction in our approach, we further introduce aesthetic dress optimization to tailor a dress for a virtual character by adjusting the clothing details. Our approach deeply mines the relationship between scene information and clothing attributes, and considers the different requirements for clothing of different people, so as to synthesizes suitable outfit for the virtual characters in different scenes that conform to his/her characteristics.

## 3 Overview

Given a scene as input, our approach aims to synthesize dress for a virtual character so that the virtual character can have an appropriate and professional look to complete interaction tasks. The scene can be captured through an RGB camera mounted on an MR device, e.g., Hololens.

We use the work-flow similar to a fashion consultant to tackle the problem synthesizing dress for a virtual character. When a fashion consultant tries to give suggestions on a dress, the dress code is obtained firstly, follow which the consultant recommends a dress accordingly. Therefore, the first step of our approach is to determine the dress code. We come up with an approach to automate the dress code prediction from a scenario. The dress code is represented as a set of dress category. After that, our approach tailors a dress for the virtual character through tuning details of the dress.

To this end, we devise a pipeline to synthesize an aesthetic dress with respect to the scene in which the virtual character will show up. The pipeline is shown in Figure 2. The pipeline comprises two phases: *Dress Code Prediction* and *Aesthetic Dress Optimization* . In the *Dress Code Prediction* phase (b), our approach first performs visual analysis of the input scene, obtaining season, occasion, and scene category. Such analysis results are used to predict "what to wear" (i.e.. dress code) to match the semantic information.

To perform an aesthetic look, the generated outfit needs to convey harmonious design details for the character-specific appearance (e.g., gender, age, height, skin color, etc). For example, tall women prefer tops of long or medium length, which give full play to their height advantage. Fat people should avoid wearing too tight to show their body defects, which improves the comfort of their clothes. In the *Aesthetic Dress Optimization* phase (c), our approach uniformly optimizes the details of the dress, such as the collar type (turtleneck, ruffle, etc), waist type of pants (high-rise, low-rise, etc). After the optimization, an aesthetic dress is output (d). By applying our approach, virtual characters could be automatically dressed in an aesthetic dress in different scenes.

## 4 Dataset

There are several existing clothing datasets. However, to the best of our knowledge, none of them is suitable to evaluate the scene-aware dress generation for virtual characters task. A large full-annotated benchmark dataset [17], employing full-body fashion show images downloaded from Vogue and containing 550 fashion brands, is released to tackle the problem of clothing fashion style understanding. However, all images are taken from shows, and all people are professional models with tall and slim figures, making it lack of rich scene information and extensive character characteristics. The Generative Fashion Dataset [22] introduce a dataset of fashion images paired with item descriptions, but it provides a uniform white background, making it unable to get any scene information. The iMaterialist Fashion Attribute Dataset [9] includes over one million annotated fashion images, labelling a total of 228 fashion attributes. However, it labels too few clothing visual features, and lacks scene information and character features.

We created a dataset, *Scene-Relevant Dress Dataset*, to achieve the goal of dress synthesis for virtual characters: i) learning the relationship between scene semantic and dress code; ii) learning relationship between VR character-specific appearance and dress details.

The dataset consists of: i) 2,993 combinations of different scenes, characters, and the corresponding dress, which were obtained from 1,200 images collected from professional photography websites and a human actions dataset [35]. These data are considered as aesthetic data, i.e. *positive examples ("Good")*, since the images come from real photos, in which people are dressed suitably in a specific environment. ii) 2,400 *negative examples ("Bad")* synthesized by randomly sampling different scenes, character appearance, and dress.

Moreover, we invited annotators who are well trained with the annotation method to further annotate each image in our *Scene-Relevant Dress Dataset* (as shown in Figure 3). The annotators were required to annotate three types of labels, which are as follows. *Please find the full annotation list in the supplementary materials.*

- *Scene Attributes*, including 4 kinds of seasons (i.e. spring, summer, autumn, and winter), 5 kinds of occasions (i.e. daily, sports, trip, party, and art photos), 80 categories (e.g., swimming pool, hospital, library, etc.), and color. Specifically, the scene category was pre-classified by a trained neural network, Inception-ResNet-V2 [25]. The scene color was pre-extracted using k-means [18].
- *Character Appearance Attributes*, including gender (i.e. male, female), age (i.e. child, adult, elder), height (short, medium, high), body shape (i.e. thin, medium, fat), skin color (i.e. white, yellow, balck), and hair color (i.e black, yellow, white), which are demonstrated to influence the aesthetic expression with their dress [6, 28].

Figure 3: Annotation examples in our dataset. In each group, the left column is the example image, and the right is the annotation of scene attributes (first row), character appearance attributes (second row), and dress attributes (third row).

- *Dress Attributes*, including dress types (e.g., top, skirts, and romper), categories (e.g., sportswear, down jackets), materials (e.g., wool, chiffon), and detail geometry features (e.g., collar shape and cuff shape for the top, and waist type for the bottom).

Moreover, for negative examples ("Bad"), we synthesized about 3,600 samples by randomly sampling different scenes, character appearance, and dress, and selected 2,400 negative samples by evaluating the compatibility of these attributes. We recruited three participants to evaluate the compatibility of these randomly generated attributes on scenes, characters, and dresses. The rating ranges from 1 to 5, i.e. a 1-5 Liking scale, with 1 meaning that dress is in violation of scenes and characters, and 5 meaning the opposite. Finally, we chose 2,400 samples with the rating below 3 as a negative example. Please refer to supplementary material for the detailed label of each annotation.

## 5 Dress Code Prediction

In our approach, the type of dress is affected by three scene features: season [23], occasion [15], and category [24]. These three features have been proven to be strongly associated with high-level scene semantics (i.e. style expression) [38]. Inspired by these studies, our approach is designed to first predict the dress type for the input scene such that the synthesized dress matches the scene in terms of the expressed attributes (i.e. season, occasion, and category).

We trained a fully connected neural network as a binary classifier to predict whether a dress code is suitable. We define the dress code as the determination of dress type and the corresponding material. With the input of different combinations of and material, scene features, and character gender, the classifier outputs whether the dress type and materials are reasonable, i.e. we could obtain the appropriate combination of dress type and corresponding materials. Because the number of combinations is limited, we use brute-force search to get the solution.

Assume the a virtual character can be dressed 5 items at most. We present a dress code as a $5*2$ matrix $\mathbf{c} = [\mathbf{t}, \mathbf{m}]$. $\mathbf{t} = [t_1, t_2, t_3, t_4, t_5]^T$ depicts the dress type. For example, $t_i$ is the $i$-th item's type. It is worth noting that if the item does not exit, the value will be set as 0. $\mathbf{m} = [m_1, m_2, m_3, m_4, m_5]^T$ depicts the corresponding dress material. To predict the dress code $\mathbf{c}$, we train a fully connected neural network. The network consists of three fully layers (256, 128, and 2 output sizes). The first two layers are applied with ReLU activation, and the last layer is applied with Sigmoid activation.

**Training Data.** To train our model, we randomly selecte 2,400 combinations of scenes and corresponding character for the collected *Scene-Relevant Dress Dataset*; the rest were used for evaluation. Specifically, the scene attributes are encoded as 3D features, i.e. representing season, occasion, and category, which was concatenated with the 1D feature of the character's gender. Each combination data are annotated with a 10D-feature label, i.e clothing type (top, pant, skirt, dress, and romper) and corresponding materials, as described in Sec. 4.

**Learning.** We use Binary Cross Entropy as the loss function. In each learning task, 80% of the examples were fed into the network, and 80% were for training and 20% were for testing in each epoch. We set the batch size of the training network to 32 and the momentum to 0.9. The learning rate is initialized to 0.0001, and the weight attenuation is 0.1. We adjust the learning rate every 50 epoch.

**Evaluation.** We test the performance of our model on the 20% test set of the collected *Scene-Relevant Dress Dataset*. The accuracy of a single evaluation of style classifier is 95.67%. We also compared the performance of our model against different architectures, such as SVM, achieving a prediction accuracy of 57.14%, and find that the fully connected network achieves the best results.

## 6 Aesthetic Dress Optimization

Based on the dress type predicted from the scene image and VR character gender, our approach further refines the dress synthesis in terms of the detail geometries to yield attractive dress for the virtual character. Incompatible dress detail design to the scene has been shown to influence one's dress expression [37]. To address the challenge of aesthetic expression, we formulate the dress detail geometry synthesis process as optimization with constrains of aesthetic expressions. We will discuss how we define such constraints or cost terms in detail.

### 6.1 Cost Function

As demonstated in [15], the reasonable dress is reflected through properly wearing and aesthetically wearing [33] for the specific scene. This inspires us to define our cost function from the aspects of the dress code (i.e. dress types and the corresponding materials) and detailed aesthetics. We synthesize dress by optimizing against a defined total cost function (Equation 1). The configuration of a dress is defined as $\Theta = \{\theta_i\}, i \in [1, 13]$. Each dimension $\{\theta_i\}$ represents a certain detail geometry feature. For example, $\theta_2$ represents collar shape of the top. We solve the problem of searching for an optimal configuration $\Theta^\star$ by optimizing a cost function:

$$\mathrm{C}(\Theta, \Phi, \Psi) = \mathrm{C_s}(\Theta, \Phi, \Psi) + \lambda \mathrm{C_o}(\Theta), \tag{1}$$

where $\Phi = (\varphi_{season}, \varphi_{occasion}, \varphi_{category})$ represents the input scene attributes, $\Psi = (\psi_{gender}, \psi_{age}, \psi_{height}, \psi_{shape})$ represents the VR character appearance. $\mathrm{C_s}(\cdot)$ is a single item cost for evaluating the aesthetic expression between each item (e.g., top, bottom) and the input (i.e. the scene and the VR character). $\mathrm{C_o}(\cdot)$ is a whole cost, which measures the aesthetic expression among dress items. $\lambda$ is a weight to balance these two terms, which is set to be 0.5 as default.

**Aesthetic Evaluation Model.** Similar to the *Dress Code Prediction* model, we train two same-architecture fully connected neural networks for $\mathrm{C_s}(\cdot)$ and $\mathrm{C_o}(\cdot)$, respectively. Taking the evaluation model of a single item as an example, we trained the model with our *Scene-Relevant Dress Dataset*, which contains 1,200 different scenes and the corresponding single dress item annotated with aesthetic expression (i.e. positive/"Good" or negative/"Bad"). We test the performance of our two models on the 20% test set of the collected *Scene-Relevant Dress Dataset*. The accuracy of a single evaluation of the single clothing item classifier is 91.29%, and that of the whole dress classifier is 90.67%.

Figure 4: Aesthetic dress synthesis results for different VR characters in the same *"Spring" "Forest"*.

$C_s(\cdot)$ drives each synthesized dress item in $\mathbf{c}$, dressed on the VR character, to perform aesthetic expression with the scene. It is defined as $C_s(\cdot) = \frac{1}{Z}\sum_{i=1}^{5} w_i p(\Phi,\Psi)$. $w_i$ is calculated by the dress type $\mathbf{t}$ in $\mathbf{c} = [\mathbf{c},\mathbf{m}]$. If $t_i = 0$, then $w_i = 0$, otherwise, $w_i = 1$. $Z$ is a normalization parameter, which equals to the number of $t_i$ where $t_i \neq 0$. $p$ is defined as:

$$p(\cdot) = 1 - \frac{exp(x_1)}{exp(x_1) + exp(x_2)}. \tag{2}$$

$[x_1, x_2]^T$ is the output of the fully connected layer, reflecting the possibility of whether the dress item matches the inputs.

Based on the model trained with 1,200 scenes and the corresponding 2,993 dresses we elaborate the cost $C_o(\cdot)$ similar to Equation 2.

### 6.2 Optimization

Motivated by the fact that there could be multiple optimal clothing for the same scene and the same person, the dress synthesis could be achieved by sampling in the solution space with a density function defined using idealized analytical formulations.

For the cost function $C(\cdot)$in Equation 1, we adopt genetic algorithm (GA) sampler to explore the configuration space iteratively. In each iteration, the sampler generates multiple candidates (i.e. multiple detail geometries $\Theta$), considered as a "group" together with the previous candidates, are retained or removed by specific rules. We use a fitness function to determine which clothing detail geometry feature to keep or eliminate:

$$\Gamma = \frac{1}{C(\Theta,\Phi,\Psi)} \tag{3}$$

The initialization of each detail chromosomes is a random sampling of each dimension. In the process of optimization, the fitness value determines the survival or elimination probability of each chromosome.

We use two strategies to propose a 'move': cross move and mutation move.

**Cross Move.** The process of splicing two existing chromosomes to generate a new chromosome. We use roulette algorithm to randomly select two parent chromosomes, $\theta_m$ and $\theta_n$. A crossing position $i$ is selected by a uniform random algorithm. The new chromosome $\theta_n$ is generated via concatenating of $\theta_m[1:i]$ and $\theta_n[i:13]$. In order to perform aesthetic expression, we enhance the probability of selecting parent chromosomes according to their calculated cost. The selection probability of each chromosome is defined as:

$$q_i = \frac{a_i}{\sum_{N\times(1-\alpha)+M}^{1} a_i} \tag{4}$$

where $a_i$ is the fitness value of chromosome $i$. This allows individuals with large fitness to have a greater probability to be selected as parents to generate new individuals. In other words, the more aesthetic the details are, the more likely they will be preserved.

**Mutation Move.** Since we treat a dress that defines the all detailed attributes of every dress item as an individual. For example, we determined that a dress is composed of a jacket and a pair of jeans in the phase of type prediction. In a certain iteration in the aesthetic dress optimization, we sampled the top as loose model, round neck and long sleeves, and the jeans as slim model and ankle shaped. Such a dress of these attributes can be regarded as an individual. A population is a collection of all such individuals. In order to improve the diversity of the population, in each iteration, we randomly select several chromosomes for mutation. For each chromosome to be mutated, we randomly select one gene and change its value, which is equivalent to fine-tuning dress, for example, changing bubble sleeves to straight sleeves. This method, called mutation, helps to improve the diversity of the population, and try not to destroy the good quality of the population.

As a result, in each iteration, the genetic algorithm calculates the individual fitness of the population. The chromosomes with higher fitness will be considered as "Good" ones, and the "Bad" chromosomes, on the contrary, will be eliminated based on an elimination rate $\alpha = 0.2$. We terminate the optimization when the average difference of population fitness between the two iterations is less than 5% over the past 10 iterations.

### 6.3 Coloring

When the dress detail features are determined through the optimization process, our approach automatically coloring each dress item by a color palette, which is selected by the following steps:

- apply the method proposed by [18] to extract input scene color palette. The palette size is set as 5;
- apply the algorithm proposed by [20] to calculate the harmony between the scene color palette and each color palettes provided by Adobe Kuler and COLOURLovers.
- based on the dress design guidelines [38], we coloring the synthesized dress by color saturation.

In practice, we find that this leads to a more comfortable experience for evaluating dress aesthetic expression.

## 7 Experiment

We implemented our approach using Phython 3.6 and Marvelous Designer 9 Enterprise. We ran our experiments on a PC equipped with 16GB of RAM, a Nvidia Titan X graphics card with 12GB of memory, and a 2.60GHz Intel i7-5820K processor. Once the dress detail features are determined, dress synthesis will be completed automatically in Marvelous Designer 9 Enterprise. We build a dress patch database, including sleeves, cuffs, front and so on. Each patch is equipped with different attributes. For example, the sleeve includes long sleeve, medium sleeve and short sleeve; the front includes zipper, pullover and open. According to the given dress details, our system automatically selects the patches and stitches the corresponding patches. Considering different body shapes of


Figure 5: Different results of Random Synthesis, our approach, and Professional Synthesis that used in our experiments.

virtual characters, the patch will be resized according to the body parameters, which are automatically provided by the engine, so as to ensure the dress perfectly fits the character.

We applied our approach for synthesizing aesthetic dresses to different scenes and different VR character. The scenes cover different seasons, occasions, and categories, including seaside restaurant for parties in spring, playground for sports events in summer, meeting room for conferences in autumn, and cabin for vacations in winter, etc. Please refer to supplementary materials for all synthesized results, which were also used in our perceptual study (Section 7).

When applying our approach, the user can synthesize aesthetic dresses consistent with the scene attributes and the VR characters' appearance. In addition, we demonstrate how our approach can be applied to tackle several common scenarios in aesthetic dress synthesis:

- **Same Scene with Different VR Characters.** Given the same scene, our method can synthesize different aesthetic dresses for different characters, which are tailored according to their different appearance attributes. As shown in Figure 4, our approach synthesizes 6 aesthetic dresses for VR characters under the same *"Spring" "Forest"*.
- **Different Scenes with the same VR Character.** As shown in Figure 1, the same character can be dressed in different clothes when appearing in different scenes. In this case, the adult female with white skin, yellow hair, standard and height dressed in a business suit, sportswear, etc. in different scenes.

In the following sections, we discuss several quantitative and qualitative experiments to evaluate the effectiveness of our aesthetic dress synthesis approach, which we find to be consistent with the target visual content conveyed by the scene and VR character.

### 7.1 Methods for Comparison

**Different Approaches.** We compared three approaches of aesthetic dress synthesis:

- Random synthesized aesthetic dresses;
- Our synthesized scene-aware aesthetic dresses.
- Professional Synthesis. We recruited three professional fashion designers with more than five years' experience in fashion design.

**Validation Dataset.** We used 7 scene images (covering 4 seasons, 5 occasions, and 7 categories) and 6 VR characters expressing different appearances as input, and obtain 49 aesthetic dresses for each method aforementioned. Please refer to supplementary materials for the validation dataset.

Table 1: Demographics of study participants. G=Gender (M=Male; F=Female). Occ.=Occupation (STU=Students; EDU=Educators; MER=Merchants; TEX=Technicians). Fash.Exp.=Fashion Experience (NPE=People with no fashion experience; ALS=Amateurs with low skills level; AHS=Amateurs with high skills level; PRO=Professional fashion designers).

| G | Age | Occ | Fash.Exp |
|---|---|---|---|
| M:14 | <20:3 | STU:13 | NPE:5 |
| F:11 | 20-30:18 | EDU:3 | ALS:7 |
| | 30-40:1 | MER:2 | AHS:11 |
| | 40-50:3 | TEC:7 | PRO:2 |

**Procedure.** We compared the validation dataset of these approaches in quantitative and qualitative experiments.

*Quantitative Experiment.* We measured the objective performance of the compared approaches with respect to dress aesthetic expression and synthesis time.

- Dress aesthetic expression. Due to the lack of computational methods for outfits evaluation, we use the cost function as a reference metric, i.e. for the outfits synthesized by different approaches, we compute the average aesthetic scores ($score = 1 - \mathrm{C}(\cdot)$). A higher score (a lower cost) means the synthesized outfit is regarded as "Good" from the perspective of classifiers.
- Synthesis time. We recorded the synthesis time of the outfit synthesis process of three different approaches.

*Qualitative Experiment.* Due to the lack of computational methods for scene-aware dress evaluation, we use the cost function as a reference metric. Since the cost function is not a "perfect" evaluation metric, we further use qualitative (subjective) evaluation to validate our approach.

We recruited 25 volunteer participants who reported normal or corrected-to-normal vision with no color-blindness. The detailed participants' demographics are shown in Table 1. As shown, the participants represent a diversity of backgrounds in terms of gender (14 men, 11 women), age (range is 18 to 50 with a mean of 24.56), occupation (from people who are students, to those who are educators, Technicians and merchants). Before each study, the participants were given a task description and encouraged to ask any questions. The participants were seated 35 cm in front of a screen (with $1440 \times 900$ resolution).

The purpose of this experiment is to evaluate the aesthetic expression of our results synthesized under the input scene and the virtual character. We asked the participants to rate the dress code rationality, single item aesthetic expression, and overall dress aesthetic expression for each data, using a 1-5 Likert scale, with 1 meaning bad aesthetic performance and 5 meaning the opposite. The dresses are randomly selected from our validation dataset of different approaches to avoid bias.

### 7.2 Results and Discussion

**Quantitative Experiment.** We conducted a quantitative experiment with each approach discussed in Sec. 7.1, synthesizing aesthetic dresses for VR characters in different scenes. For dress aesthetic expression, our approach attained the highest score (M = 0.97, SD = 0.01), closely followed by professional approach (M = 0.94, SD = 0.02). The strategy of training the dress aesthetic classifier on *Scene-Relevant Outfit Dataset* has a comparable capability to professional evaluation. The random synthesis results obtain the lowest average score (M = 0.63, SD = 0.22 ), which verifies the synthesized outfit without any constrains can not satisfy the requirements of different scenes in most cases.

For each scene and the corresponding VR character, we recorded the synthesis time of the aesthetic dress synthesis process for each

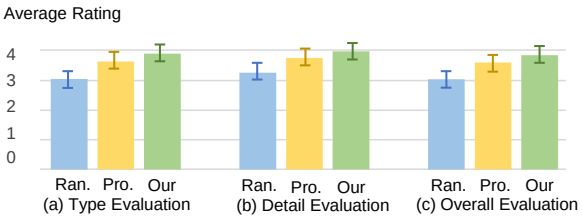


Figure 6: Average user ratings of scene-relevant outfit compatibility expression of different approaches, i.e. our approach, Random Synthesis, and Professional Synthesis.

approach. The results show that the *Random Synthesis* uses the least time to synthesize an outfit (M = 0.54 ms, SD = 0.10 ms). Our approach synthesizes a pose much faster (M = 16.86 s, SD =0.18 s) for 250 iterations in the optimization process, compared to *Professional Synthesis* (M = 48.53 mins, SD = 1.84 mins). The experts claimed that they first need to analyze the visual information conveyed by both the scene and the presented virtual character in 10 minutes. Then, they decide what to wear according to the season and the occasion. Designing details for clothes is the most time-consuming process, which cost at least 20 minutes.

**Qualitative Experiment.** Our approach obtained the highest rating in all four evaluations (dress code: M = 3.92, SD = 1.01; single item: M = 4.00, SD = 0.94; overall: M = 3.87, SD = 0.92), followed by *Professional Synthesis* (dress code: M = 3.65, SD = 1.15; single item: M = 3.78,SD = 1.08; overall: M = 3.62, SD = 1.07) and *Random Synthesis* (dress code: M = 3.28, SD = 1.50; single item: M = 3.07,SD = 1.30; overall: M = 3.06, SD = 1.34). In all cases, our syntheses are more preferable than random syntheses and are comparable to professional syntheses.

To ascertain that our results are efficacious, we performed the OneWay ANOVA test in qualitative experiments, using a significance level of 0.05. Our hypothesis $H_0$ was that users believe that the performances of dressed synthesized by different approach are similar, i.e. ratings for our approach and *Random Synthesis* are similar, and ratings for our approach and *Professional Synthesis* are similar. The results suggest that our approach obtained higher score than Random Synthesis significantly: dress code ($F_{[1,108]} = 4.727$, p = .003 $< .05$, $\eta_p^2 = 0.317$); single item ($F_{[1,108]} = 3.916$, p = .008 $< .05$, $\eta_p^2 = 0.304$); overall ($F_{[1,108]} = 4.841$, p = .003 $< .05$, $\eta_p^2 = 0.333$). In all cases, we have p-values less than 0.05. Therefore, we reject the null hypothesis $H_0$ in all categories. On the other hand, comparison results show that the ratings between our approach and *Professional Synthesis* are mostly comparable, i.e. dress code ($F_{[1,108]} = 2.264$, p = .143 $> .05$, $\eta_p^2 = 0.122$); single item ($F_{[1,108]} = .612$, p = .663 $> .05$, $\eta_p^2 = 0.110$; overall ($F_{[1,108]} = 0.716$, p = 0.515 $> .05$, $\eta_p^2 = 0.124$). It turns out that our approach can synthesize aesthetic dresses for the virtual character in different scenes.

To verify the factors considered in our optimization contributed to the overall aesthetic expression, we computed Bivariate (Pearson) correlation coefficients between the ratings of the overall expression and other factors, respectively. Our hypothesis $H_0$ was that there is no correlation present between the overall expression and the dress code, or between the overall expression and the single item aesthetic expression. There are positive correlations between the overall expression and the dress code (r = .651, p $< .05$); between the overall expression and the single item aesthetic expression (r = .757, p $< .05$); In all cases, we have p-values less than 0.05. Therefore, we reject the null hypothesis $H_0$ in all experiments. The higher performance the type and detail are, the higher the overall experience.

**User Feedback.** Most users commented that they could feel the dress changing accordingly to the requirements of the scene.


Figure 7: Results of different interactions.

However, some participants commented that the dress material is not real, especially the material of sweater and chiffon, which is caused by the lack of simulation authenticity. Moreover, some users commented that the absent expression of characters would affect the overall judgement. This can be explained by the fact that fashion aesthetics perception of different users is inconsistency.

## 8 Conclusion

In this paper, we propose a computational approach for synthesizing dress for virtual characters automatically. Considering the scene semantic information and the virtual character's attributes, our approach synthesizes dress that satisfies the functionality; meets the aesthetic criteria; and fits the virtual character's attributes. Perceptual studies also find that the synthesized dress is reasonable and consistent with the application scenario.

Our approach leads to a variety of potential applications. For example, our approach could handle the dress design requirements of the large amount of the virtual anchor, which has been used in live streaming. The character for selling products (as shown in Figure 8) could dress in our synthesized clothes that consistent with the live streaming room theme, so as to enrich the shopping experience for customers. With MR devices, the virtual anchor could be visualized in real stores, which could change his/her clothes every day according to different places, so as yield to more natural and diverse interaction experience for customers. Moreover, our approach could also be applied with gaming techniques to synthesize aesthetic dresses for a large number of characters with less human effort. In addition, our approach can be applied to real-world applications with real-world outfits. For example, our approach could provide reasonable outfit references for dress designers, or assist the clothing recommendation system [36] to help users dress appropriately to the scene environment.

*Limitation and Future Work.* As the performance is limited to the classifier trained on the collected dataset, some of the inputs may get unexpected results. Figure 7 shows that dress requirements differ even in the same scene. For example, given a specific scene as shown (season: summer, occasion: sports, scene category: lake), the virtual character's dress changes due to different actions. Due to the small granularity of occasions, we get the result (a). In reality, the occasion "sports" can be subdivided. The virtual character in the scene may run on the track or swim in the lake. Thanks to the strong flexibility of this framework, we get result (b) by adding interaction constrains, specifying the action as swimming, which proves the scalability of our approach.

As an initial attempt, we have introduced three aspects of scene semantics (e.g., season, category and occasion), which can be enriched and more specific. For example, weather and temperature will help refine the synthesis of dress materials. Events and behaviors of people enrich the semantic information of the occasion, thus optimizing the selection of dress types. In our approach, we define character appearance attributes discretely, but many of the


Figure 8: Aesthetic dress synthesis for the virtual anchor in live streaming of product selling. The synthesized dress in harmony with the virtual store and the character's appearance with aesthetic pose [29] could enhance the user experience.

character's appearance attributes are continuous, such as age, height, body, etc. We believe that more specific attribute annotation will help to achieve more diverse and reasonable outfits.

For simplicity, we synthesized the dress color with monotonous colors referring to a popular color scheme. Creative color stitching, texture and pattern may help to improve the synthesis further. Also, a comprehensive study of coloring the clothing item is necessary, e.g., dress codes sometimes constraint the color of the items.

## REFERENCES


[1] E. D. Aguiar, L. Sigal, Y. A. Treuille, and J. K. Hodgins. Stable spaces for real-time clothing. *ACM Transactions on Graphics*, 29(4), 2010.

[2] K. E. Ak, J. H. Lim, J. Y. Tham, and A. Kassim. Semantically consistent hierarchical text to fashion image synthesis with an enhanced-attentional generative adversarial network. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3121–3124, 2019. doi: 10.1109/ICCVW.2019.00379

[3] D. Baraff and A. Witkin. Large steps in cloth simulation. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98. Association for Computing Machinery, New York, NY, USA, 1998. doi: 10.1145/280814.280821

[4] L. Chen, J. Tian, G. Li, C.-H. Wu, E.-K. King, K.-T. Chen, S.-H. Hsieh, and C. Xu. Tailorgan: Making user-defined fashion designs, 2020.

[5] S. Dobbyn, R. McDonnell, L. Kavan, S. Collins, and C. O'Sullivan. Clothing the Masses: Real-Time Clothed Crowds With Variation. In D. Fellner and C. Hansen, eds., *EG Short Papers*. The Eurographics Association, 2006. doi: 10.2312/egs.20061038

[6] J. Fan, W. Yu, and L. Hunter. *Clothing appearance and fit: Science and technology*. Elsevier, 2004.

[7] W.-W. Feng, Y. Yu, and B.-U. Kim. A deformation transformer for real-time cloth animation. *ACM Trans. Graph.*, 29(4), July 2010. doi: 10.1145/1778765.1778845

[8] P. Guan, L. Reiss, D. A. Hirshberg, A. Weiss, and M. J. Black. Drape: Dressing any person. *ACM Transactions on Graphics*, 31(4):1–10, 2012.

[9] S. Guo, W. Huang, X. Zhang, P. Srikhanta, Y. Cui, Y. Li, S. Belongie, H. Adam, and M. Scott. The imaterialist fashion attribute dataset, 06 2019.

[10] X. Han, Z. Wu, Y.-G. Jiang, and L. S. Davis. Learning fashion compatibility with bidirectional lstms. In *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1078–1086, 2017.

[11] Y. Hu, X. Yi, and L. S. Davis. Collaborative fashion recommendation: A functional tensor factorization approach. In *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 129–138, 2015.

[12] V. Jagadeesh, R. Piramuthu, A. Bhardwaj, W. Di, and N. Sundaresan. Large scale visual recommendations from street fashion images. *CoRR*, abs/1401.1778, 2014.

[13] J. M. Kaldor, D. L. James, and S. Marschner. Simulating knitted cloth at the yarn level. *Acm Transactions on Graphics*, 27(3):1–9, 2008.

[14] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke. Improving outfit recommendation with co-supervision of fashion generation. In *The World Wide Web Conference on*, pp. 1095–1105, 2019.

[15] S. Liu, T. V. Nguyen, J. Feng, M. Wang, and S. Yan. Hi, magic closet, tell me what to wear! In *ACM Multimedia*, 2012.

[16] X. Liu, J. Li, J. Wang, and Z. Liu. Mmfashion: An open-source toolbox for visual fashion analysis. *arXiv preprint arXiv:2005.08847*, 2020.

[17] Y. Ma, J. Jia, S. Zhou, J. Fu, Y. Liu, and Z. Tong. Towards better understanding the clothing fashion styles: A multimodal deep learning approach. In *Thirty-First AAAI Conference on Artificial Intelligence*, pp. 38–44, 2017.

[18] J. Macqueen. Some methods for classification and analysis of multivariate observations. *Proc.symp.math.statist. & Probability*, 1, 1967.

[19] T. Nakamura and R. Goto. Outfit generation and style extraction via bidirectional lstm and autoencoder. 2018.

[20] Peter, O'Donovan, Aseem, Agarwala, Aaron, and Hertzmann. Color compatibility from large datasets. *Acm Transactions on Graphics*, 2011.

[21] Q. Ping, B. Wu, W. Ding, and J. Yuan. Fashion-attgan: Attribute-aware fashion editing with multi-objective gan, 2019.

[22] N. Rostamzadeh, S. Hosseini, T. Boquet, W. Stokowiec, Y. Zhang, C. Jauvin, and C. Pal. Fashion-gen: The generative fashion dataset and challenge, 2018.

[23] F. Salata, I. Golasi, V. Ciancio, and F. Rosso. Dressed for the season: Clothing and outdoor thermal comfort in the mediterranean population. *Building and Environment*, 146:50–63, 2018.

[24] E. Simo-Serra, S. Fidler, F. Moreno-Noguer, and R. Urtasun. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 869–877, 2015.

[25] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016.

[26] H. Tsujita, K. Tsukada, K. Kambara, and I. Siio. Complete fashion coordinator: a support system for capturing and selecting daily clothes with social networks. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 127–132, 2010.

[27] N. Umetani, D. M. Kaufman, T. Igarashi, and E. Grinspun. Sensitive couture for interactive garment modeling and editing. *ACM Transactions on Graphics (TOG)*, 2011.

[28] L. C. Wang, X. Y. Zeng, L. Koehl, and Y. Chen. Intelligent fashion recommender system: Fuzzy logic in personalized garment design. *IEEE Transactions on Human-Machine Systems*, 45(1):95–109, 2015. doi: 10.1109/THMS.2014.2364398

[29] Y. Wang, H. Sifan, N. Bing, and L. Wei. Photo stand-out: Photography with virtual character. In *ACM Multimedia*, vol. 38, 2020.

[30] K. Wolff and O. Sorkine-Hornung. Wallpaper pattern alignment along garment seams. *ACM Transactions on Graphics (TOG)*, 2019.

[31] S.-K. Wong. Modeling and simulation techniques for garments. In *Computer Technology for Textiles and Apparel*, pp. 173–199. Elsevier, 2011.

[32] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Retrieving similar styles to parse clothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(5):1028–1040, 2015.

[33] S. Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *Computer Vision & Pattern Recognition*, 2012.

[34] W. Yang, P. Luo, and L. Lin. Clothing co-parsing by joint image segmentation and labeling. In *CVPR '14 Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3182–3189, 2014.

[35] B. Yao, X. Jiang, A. Khosla, A. Lin, L. Guibas, and F. F. Li. Human action recognition by learning bases of action attributes and parts. pp. 1331–1338, 11 2011. doi: 10.1109/ICCV.2011.6126386

[36] L.-F. Yu, S.-K. Yeung, D. Terzopoulos, and T. F. Chan. Dressup!: outfit synthesis through automatic optimization. *international conference on computer graphics and interactive techniques*, 31(6):134, 2012.

[37] W. Yu, H. Zhang, X. He, X. Chen, L. Xiong, and Z. Qin. Aesthetic-based clothing recommendation. In *WWW '18 Proceedings of the 2018 World Wide Web Conference*, pp. 649–658, 2018.

[38] X. Zhang, J. Jia, K. Gao, Y. Zhang, D. Zhang, J. Li, and Q. Tian. Trip outfits advisor: Location-oriented clothing recommendation. *IEEE Transactions on Multimedia*, 19(11):2533–2544, 2017.