Virtual Try On

A Project report submitted in partial fulfilment of the requirements of the award of the degree of

Bachelor of Technology

in

Computer Engineering

by

Tripti Agrawal, Reg No: PCE20CS187
Shitanshu Goyal, Reg No: PCE20CS173
Vikram Jangir, Reg No: PCE20CS192
Shivam Patil, Reg No: PCE20CS176

under the guidance of

Prof. Barkha Narang, Assistant Professor

(Session 2021-22)

Department of Computer Engineering

Poornima College of Engineering

ISI-6, RIICO Institutional Area, Sitapura, Jaipur – 302022

December, 2021

Department Certificate

This is to certify that Ms. Tripti Agrawal, Mr. Shitanshu Goyal, Mr. Vikram Jangir, Mr. Shivam Patil, registration no. PCE20CS187, PCE20CS173, PCE20CS192, PCE20CS176, of the Second Year Department of Computer Engineering, has submitted this project report entitled “VIRTUAL TRY ON ” under the supervision of Prof. Barkha Narang, working as Assistant professor in department of Computer Engineering as per the requirements of the Bachelor of Technology program of Poornima College of Engineering, Jaipur.

Dr. Surendra Kumar Yadav	Ms. Sonam Gour
Head, Dept. of Computer Engineering	Coordinator-Project

CANDIDATE'S DECLARATION

I hereby declare that the work which is being presented in this project report entitled "Virtual Try On" in the partial fulfilment for the award of the Degree of Bachelor of Technology in (Computer Engineering), submitted in the Department of Computer Engineering, Poornima College of Engineering, Jaipur, is an authentic record of my own work done during the period from July 2021 to Dec 2021 under the supervision and guidance of Prof. Barkha Narang.

I have not submitted the matter embodied in this project report for the award of any other degree.

Signature	Signature
Name of Candidate: Tripti Agrawal Registration no.: PCE20CS187	Name of Candidate: Vikram Jangir Registration no.: PCE20CS192
Signature	Signature
Name of Candidate: Shitanshu Goyal Registration no.: PCE20CS173	Name of Candidate: Shivam Patil Registration no.: PCE20CS176

Dated: - 30-12-2021

Place: Jaipur

SUPERVISOR'S CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Signature)

Dated: (Name of Supervisor)

Place: Jaipur (Designation & Affiliation)

ACKNOWLEDGEMENT

I would like to convey my profound sense of reverence and admiration to my supervisor **Ms. Barkha Narang,** Assistant Professor **Department of Computer Engineering, Poornima College of Engineering,** for her intense concern, attention, priceless direction, guidance and encouragement throughout this research work.

I am grateful to **Dr. Mahesh Bundele**, Director of Poornima College of Engineering for his helping attitude with a keen interest in completing this dissertation in time.

I extend my heartiest gratitude to all the teachers, who extended their cooperation to steer the topic towards its successful completion. I am also thankful to non-teaching staff of the department to support in preparation of this dissertation work.

I am grateful to **Dr. Mahesh Bundele**, Director of Poornima College of Engineering for his helping attitude with a keen interest in completing this dissertation in time.

I extend my heartiest gratitude to all the teachers, who extended their cooperation to steer the topic towards its successful completion. I am also thankful to non-teaching staff of the department to support in preparation of this dissertation work.

My special heartfelt gratitude goes to **Dr. Surendra Kumar Yadav, HOD,Department of Computer Engineering, Ms. Sonam Gour, Project Coordinator, Department of Computer Engineering, Poornima College of Engineering**, for unvarying support, guidance and motivation during the course of this research.

I would like to express my deep sense of gratitude towards management of Poornima College of Engineering including **Dr. S. M. Seth**, Chairman Emeritus, Poornima Group and former Director NIH, Roorkee, **Shri Shashikant Singhi**, Chairman, Poornima Group, **Mr. M. K. M. Shah**, Director Admin & Finance, Poornima Group and **Ar. Rahul Singhi**, Director Poornima Group for establishment of institute and providing facilties my studies.

I would like to take the opportunity of expressing my thanks to all faculty members of the Department, for their kind support, technical guidance, and inspiration throughout the course.

I am deeply thankful to my parents and all other family members for their blessings and inspiration. At last but not least I would like to give special thanks to God who enabled me to complete my dissertation on time.

Tripti Agrawal, Department of Computer Engineering, 20\CS\54
Shitanshu Goyal, Department of Computer Engineering, 20\CS\33
Vikram Jangir, Department of Computer Engineering, 20\CS\59
Shivam Patil, Department of Computer Engineering, 20\CS\37

LIST OF TABLES

LIST OF ACRONYMS

Serial Number	ACRONYM	FULL FORM
1	**VTO**	Virtual Try On
2	**ACGPN**	Adaptive content generating and preserving network
3	**GMM**	Geometric Matching Module
4	**DIOR**	Dressing in Order
5	**WPF**	Windows presentation foundation
6	**VITON**	Virtual tryon Network
7	**GIC**	Grid Interval Consistency
8	**TOM**	Try on Module
9	**SGM**	Semantic Generation Module

TABLE OF CONTENT

ABSTRACT

Trying clothes in clothing stores is usually a time-consuming activity. Besides, it might not even be possible to try-on clothes in such cases as online shopping. Our motivation here is to increase the time efficiency and improve the accessibility of clothes by creating a virtual dressing room environment. The virtual try-on system uses a unique capture device and artificial intelligence (AI)-driven way to digitize items of clothing. In this project Virtual Try-On new AR and VR technologies have been developed, which form the basis for a realistic, three-dimensional, (real-time) simulation and visualization of individualized garments put on by virtual counterparts of real customers. To provide this cloning and dressing of people in VR, a complete process chain is being built up starting with the touch-less 2-dimensional scanning of the human body up to a photo-realistic 2dimensional presentation of the virtual customer dressed in the chosen pieces of clothing. The emerging platform for interactive selection and configuration of virtual garments, the "virtual shop", will be accessible in real fashion boutiques as well as over the internet, thereby supplementing the conventional distribution channels.

virtual try-on can be classified as 3D model-based approaches and 2D image-based ones where the latter can be further categorized based on whether to keep the posture or not Author talks about the technology named as Adaptive Content Generating and Preserving Network which is used to preserve the character of clothes as well as the posture of the person as to show the précised image.

CHAPTER 1

INTRODUCTION

Due to the rapid growth of technology development, our daily life is heavily affected by smart a system that facilitates our activities. For instance, online shopping grew up very fast. People are getting more used to online shopping, online auctions, etc., to purchase their interested products. This way of the transaction has become the main trend and it brings great convenience to customers. However, an issue with buying clothes online is that clients cannot try the product before they get that product.

Therefore, there is an increasing demand to develop virtual dressing rooms to simulate the visualization of dressing. Therefore, most of the researchers in previous works are taking the approach to map a 2D texture to the user's body, and build an Avatar (model). In this work, we introduce a virtual dressing room application using a Microsoft Kinect sensor. Our proposed approach is mainly based on the extraction of the user from the video stream, the alignment of models, and skin color detection. We use the modules for locations of the joints for positioning, scaling, and rotation in order to align the 2D cloth models with the user. Then, we apply skin color detection on the video to handle the unwanted occlusions of the user and the model. Finally, the model is superimposed on the user in real-time. To preserve the character of clothes as well as the posture of personas to show the précised image ACGPN technology is used. ACGPN generally involves three major modules: -

1. A semantic layout generation
2. Clothes warping module
3. An inpainting module for content fusion.

Dressing in Order (Dior), which supports 2D pose transfer, virtual try-on, and several fashion editing tasks. The key to Dior is a novel recurrent generation pipeline to sequentially put garments on a person so that trying on the same garments in different orders will result in different looks. Our system can produce dressing effects not achievable by existing work, including different interactions of garments (e.g., wearing a top tucked into the bottom or over it), as well as layering of multiple garments of the same type. (e.g., jacket over shirt over a t-shirt)

CHAPTER 2

PROBLEM STATEMENT & OBJECTIVE

Technology has weaved its way into every sphere of life and shopping is no exception. We use technology to compare and ultimately decide what they will buy. Virtually trying on items allows customers to know exactly how something looks on them. Whether it is shoes or a piece of jewellery, virtually trying on products enables customers and retailers to connect without the barrier of customers coming to a physical store for shopping. Developing a technology that will show a précised image of a person by wrapping with that product is our objective. Virtual try-on applications allow users to watch themselves wearing different clothes without the effort of changing them physically. This helps users to make quick buying decisions and, thus, improves the sales efficiency of retailers.

CHAPTER 3

LITERATURE REVIEW

Paper 01:

Towards Photo-Realistic Virtual Try-on by adaptively Generating Preservation Image content (2020)

By: Ruimao Zhang, Xiaobo Guo, Wei Liu, Wangmeng Zuo, Ping Luo

Summary-

In this paper, the author talks about the technology named Adaptive Content Generating and Preserving Network which is used to preserve the character of clothes as well as the posture of personas to show the précised image. Virtual try-on can be classified as 3D model-based approaches and 2D image-based ones where the latter can be further categorized based on whether to keep the posture or not. Most existing try-on methods focus on the task of keeping the posture and identity. This author introduces the technology named Adaptive Content Generating and Preserving Network. The proposed ACGPN is composed of three modules.

First, the Semantic Generation Module (SGM**)** progressively generates the mask of the body parts and the mask of the warped clothing regions via semantic segmentation, yielding semantic alignment of the spatial layout. Second, the Clothes Warping Module (CWM) is designed to warp target clothing image according to the warped clothing mask, where the author introduces a second-order difference constraint on Thin-Plate Spline (TPS) to produce geometric matching yet character retentive clothing images Nontarget body part composition is able to handle different scenarios flexibly in a try-on task while mask in painting fully exploits the layout adaptation ability of the

ACGPN when dealing with the images from easy, medium, and hard levels of difficulty.

Third, Content Fusion Module (CFM) going beyond semantic alignment and character retention, remains a great challenge to realize layout adaptation on visual try-on tasks. To this end, both the target clothing region is required to clearly render, and fine-scale details of body parts (i.e. finger gaps) Grid Image Comparison of STN warping results with and without the secondorder difference constraint. Preserved. Existing methods usually adopt the coarse body shape as a cue to generate the final try-on images, and fail to reconstruct fine detail.

To illustrate the limitations of existing visual try-on methods, the author divides the VITON dataset into three sub-sets of difficulty levels according to the human pose in 2D reference images. An easy sample from the VITON dataset where the person in the image is represented with a standard posture, i.e., face forward and hands down. In such a case, the methods only need to align the semantic regions between the reference and target images. Some pioneering synthesized-based methods belong to this category. The image with medium-level difficulty is generally with torso posture changes. Such a goal is usually attained by developing advanced warping algorithms to match the reference image with clothes deformation.

Paper 02:

Dressing in Order: Recurrent Person Image Generation for Pose Transfer, Virtual Try-on, and Outfit Editing

Aiyu Cui Daniel McKee the Svetlana Lazebnik University of Illinois at UrbanaChampaign {aiyucui2, dbmckee2, slazebni}@illinois.ed

Summary-

In this paper, the author proposed the framework named as Dior which stands for dressing in order which supports 2D pose transfer, virtual try-on, and several fashion editing tasks.

This system can produce dressing effects that are not possible by previous systems. DiOr sequentially puts garments on a person, so that trying on the same garments in different orders will result in different looks. DiOr scans or encode the shape and texture of garment and edits the elements separately. Existing try-on methods start by producing a mutually exclusive garment segmentation map and then generate the whole. This can only achieve one look for a given set of garments, and the interaction of garments is determined by the model. By contrast, this system incorporates a novel recurrent generation module to produce different looks depending on the order of putting on garments.

The author also introduces their encoding of garments into 2D shape and texture, enabling each to be edited separately. The shape is encoded using soft masks that can additionally capture transparency. A flow field estimation component at encoding time allows for a more accurate deformation of the garments to fit the target pose. Author introduces a recurrent generation scheme that does not rely on garment labels and can handle a variable number of garments. Author also talks about a training approach, which combines pose transfer within painting to enable the preservation of fine details. Author also mentions the experimental results of their system. Pose transfer requires changing the pose of a given person while keeping that. Fashion editing. Fashion learns to minimally edit an outfit to make it more fashionable, but there is no way for the user to control the changes. our model allows users to edit what they want by making garment selections, and changing the order of garments in a semantic manner. e plan to work on improving the quality of our output through more advanced warping and higherresolution training and generation.

Paper 03:

Virtual Dressing Room Application: Aladdin Masri, Muhannad Al-Jabi, Computer Engineering Department An-Najah National University Nablus, Palestine

Summary-

In this review paper for the recent example of virtual fitting rooms and supporting technologies was presented. A virtual dressing room application using Microsoft Kinect sensor. The proposed approach is mainly based on extraction of the user from the video stream, alignment of models and skin color detection. There is a use of the modules for locations of the joints for positioning, scaling and rotation in order to align the 2D cloth models with the user. Then, skin color detection on video to handle the unwanted occlusions of the user and the model was applied. Finally, the model is superimposed on the user in real time. The problem is simply the alignment of the user and the cloth models with accurate position, scale, rotation and ordering. First, detection of the user and the body parts is one of the main steps of the problem. In literature, several approaches are proposed for body part detection, skeletal tracking and posture estimation, and superimposing it onto a virtual environment in the user interface. Kinect driver's middleware are used for various fundamental functions and for the tracking process in combination with Microsoft Kinect.
Because of the increasing importance of Microsoft Kinect image sensor in the market, we used it and WFP to capture the user physical measurements which are as follows:-

- *Kinect General components*: The components of Kinect for Windows are mainly the following:

    - Kinect hardware: including the Kinect sensor
    - USB hub, through which the sensor is connected to the computer
    - Microsoft Kinect drivers: Windows 8 drivers for the Kinect sensor

- *WPF Application*: The Windows Presentation Foundation, is UI framework to create applications with a rich user experience. It is part of the .NET framework 3.0 and higher.

- *Microsoft Kinect SDK version 1.0*: It supports up to four Kinect sensors on a single computer, skeletal tracking, a Near Mode feature that lets the camera recognize objects just 40cm away, stability and audio, Drivers and runtime .The following diagram shows the components of the Kinect for Windows developers Toolkit.

The application keep calculating distances to continue body movement, to keep this some steps are followed which are:-

1) Convert Data - Units of distance calculated using joint -joint coordinates Results in meters so we have to convert it to pixel Data.

2) Create new joint positions - we face many problems in key points that the sensor found so we decided that we could make new points away from the original points left or right or down.

3) 2D cloth - We divide the clothes into parts of pixels data to control the movement of the body and the cloth in upper, lower frame. 4) Interaction between human and 2D cloth-

- We take skeleton Data.
- We take Depth Data.
- We take the RGB data
- We measure the upper Frame and mapping the 2D cloth.
- We measure the Lower Frame and mapping the 2D cloth.
- We measure the Lower, upper Frame and mapping the 2D cloth for both.

Paper 04:

Virtual Dressing Implementation Using Body Image –Clothe Mapping

By: Ahmad al-Qerem Zarqa University

Summary-

In this paper an application (VDRS) that realistically reflects the look and feel is implemented. The application with the help of the customer (or the person) will carry out the process of trying the clothing fast and easily, and then selecting the best for him or her, which will, consequently, help us to take advantage of the enormous capacity provided by the science of interaction between man and computer. This, in turn, will reflect positively

On the seller in terms of selling the products in a comfortable, modern way resulting in better sales and on the customer in terms of speed and comfort in scanning through all available products offered to reach the best decision in choosing clothes. One positive service this application provides specially to females is that they do not have to enter a changing room that may be exploited by some assistants in peeping (or other criminal acts) Finally, this application results in our benefitting the great capacity provided by the science of the interaction between man and computer, and setting up the base line for the Foundation for its future development to be used in other beneficial services and systems.

The application approach is to replace the body image by garment mesh surface through garment image, based on the points taken from the front view and the other is from back. The red dotted rectangle covers the works done in the preprocessing stage, including digital image processing and human model creation. In the processing stage, the garment is reconstructed around the virtual human body.

Paper 05:

VITON: An Image-based Virtual Try-on Network

BY: Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, Larry S. Davis University of Maryland, College Park

Summary-

Recent years have witnessed the increasing demands of online shopping for fashion items. Online apparel and accessories sales in US are expected to reach 123 billion in 2022 from 72 billion in 2016. Despite the convenience online fashion shopping provides, consumers are concerned about how a particular fashion item in a product image would look on them when buying apparel online. Thus, allowing consumers to virtually try on clothes will not only enhance their shopping experience, transforming the way people shop for clothes, but also save cost for retailers. Motivated by this, various virtual fitting rooms/mirrors have been developed by different companies such as TriMirror, Fits Me, etc. We presented a virtual try-on network (VITON), which is able to transfer a clothing item in a product image to a person relying only on RGB images. A coarse sample is first generated with a multi-task encoder-decoder conditioned on a detailed clothing-agnostic person representation. The coarse results are further enhanced with a refinement network that learns the optimal composition. We conducted experiments on a newly collected dataset, and promising results are achieved both quantitatively and qualitatively. This indicates that our 2D image-based synthesis pipeline can be used as an alternative to expensive 3D based methods.

There is a large body of work on virtual try-on, mostly conducted in computer graphics. Guan et al. proposed DRAPE to simulate 2D clothing designs on 3D bodies in different shapes and poses. Hilsmann and P. Eisert retextured the garment dynamically based on a motion model for real-time visualization in a virtual mirror environment. Sekine et al. Introduced a virtual fitting system that adjusts 2D clothing images to users through inferring their body shapes with depth images. Recently, Pons-Moll et al. Utilized a multi-part 3D model of clothed bodies for clothing capture and retargeting. Yang et al. recovered a 3D mesh of the garment from a single view 2D image, which is further retargeted to other human bodies. In contrast to relying on 3D measurements to perform precise clothes simulation, in our work, we focus on synthesizing a perceptually correct photo-realistic image directly from 2D images, which is more computationally efficient. In computer vision, limited work has explored the task of virtual try-on. Recently, Jetchev and Bergmann proposed a conditional analogy GAN to swap fashion articles.

Paper 06:

LA-VITON: A Network for Looking-Attractive Virtual Try-On

by: Hyug Jae Lee, Rokkyu Lee, Minseok Kang, Myounghoon Cho, Gunhan Park NHN Corp.

Summary-

In this paper, we propose an image-based virtual try-on network, LA-VITON, which allows the generation of high-fidelity try-on images that preserves both the overall appearance and the characteristics of clothing items. The proposed network consists of two modules: Geometric Matching Module (GMM) and Try-On Module (TOM). Since GAN-based methods show their limitations in generating fine and sharp images, adoption of an operation, such as GMM, is necessary to deliver information directly from the source to the target. The coarse layers of TOM composites the warped clothes from GMM and an intermediate person-image. The following refinement layers remove the artifacts and enhance the quality of the generated image from the coarse network. The inshop clothes are warped and roughly aligned through GMM. The proposed scheme to learn matching and warping of clothes result in non-degraded and non-distorted warped results. TOM generates the intermediate person-image and builds a mask for composition of the warped in-shop clothing image and the generated intermediate person-image. Geometric Matching Module: One of main purposes of GMM is to align clothing while preserving the characteristics. In GMM, the perspective and the TPS transformation are trained together in every step. The perspective transformation has an additional train path for back-propagating.

Grid Interval Consistency: TPS transformation generally shows good performance but its high flexibility often causes distortion of patterns and prints. Clothes are deformable objects, but deformation on a human torso is very limited. Therefore, geometric matching needs to be restricted to maintain the shape and appearance of clothing. We introduce grid interval consistency (GIC) loss which retains the characteristics of clothes after warping. In

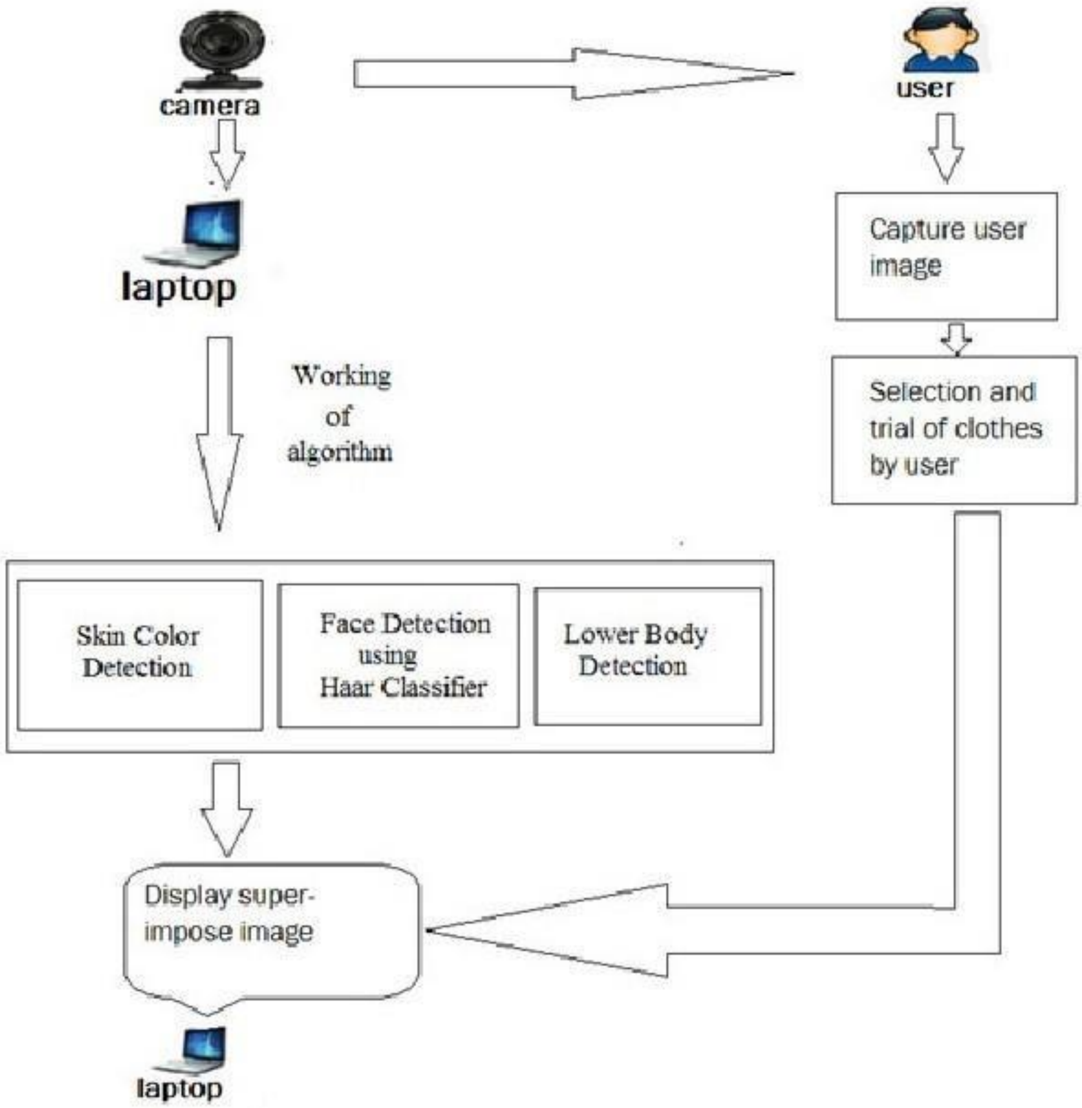
addition to patterns and prints, shapes are also preserved after warping by maintaining the consistency of the intervals.

Occlusion Handling: In real life environment, clothes on person are easily occluded by hair and arms. Since we use clothes region extracted using LIP (Look-Into-Person) as ground truth, the network tries to fit input in-shop clothes into the extracted observable clothes region. As a result, the clothes are deformed unreasonably. We address the problem by excluding the occluded regions. Thus, the network conducts accurate geometric transformation regardless of occlusion by hair and arms

Try-on Module: The purpose of TOM is to synthesize the final try-on image for the input person-image. To preserve the original details of in-shop clothing items, the warped clothing should be exploited as much as possible. If the warped clothing is directly copied onto the target person-image, unnatural seams will appear. A seamless image is obtained by blending the warped clothing from GMM.

CHAPTER 3

PROPOSED APPROACH



Comparison Table

S.No	Paper title	Author's Name	Year	Approach used	Finding	S/w and H/w Required
1	Towards PhotoRealistic Virtual Try-on by adaptively Generating Preservatio n Image content	Ruimao Zhang, Xiaobo Guo,Wei Liu,Wang meng Zuo,Ping Luo	2020	Adaptive Content Generating and Preserving Network(ACG PN)	Correct method to wrap clothes over the person body	
2	Dressing in Order: Recurrent Person Image Generation for Pose Transfer, Virtual Tryon and Outfit Editing	Aiyu Cui Daniel McKee Svetlana Lazebnik Universit y of Illinois at Urbana-Champai gn	2021	dressing in order(DiO r)	Produce dressing effects which are not possible by previous method	

3	Virtual Dressing Room Application	Aladdin Masri, Muhannad Al-Jabi, Computer Engineering Department An-Najah National University Nablus, Palestine	2019	Virtual dressing room, Microsoft Kinect, WPF.	The alignment of the user and the cloth models with accurate position, scale, rotation and Ordering.	Kinect hardware, USB hub, Microsoft Kinect driver
4.	Virtual Dressing Implementation Using Body Image –Clothe Mapping:	Ahmad al-Qerem Zarqa University	2016	Component, formatting, style, styling, insert.	VDRS application	
5	Viton: An Image based Virtual try on Network	Xintong Han,Zuxuan Wu,Davis University of Maryland	2016	RGB Image	Multi task encoder decoder generator	
6	LA-VITON: A Network for Looking Attractive Virtual try on	Hyug Jae Lee Rokkyu Lee, Minseok Kang, Myounghoon Cho, Gunhan Park NHN Corp.	2019	LAVITON, Geometric matching module, Try on module	LA -VITON produces high fidelity output images while preserving the patterns and prints.	

CHAPTER 4

CONCLUSION

The purpose of the application is to make easier the process of trying clothes while shopping, which would provide comfort for both the vendor and the customer, Reducing the time and helping people to select a wide range of clothing were a motivation to make a program that helps in this area, so it has become important (very necessary) to make the process of trying and buying of clothes more comfortable, easier and more efficient. Our motivation here is to increase the time efficiency and improve the accessibility of clothes try on by creating a virtual dressing room environment. By using ACGPN, which aims at generating photo-realistic try-on result while preserving both the character of clothes and details of human. The results clearly show the great superiority of ACGPN over the state-of-the-art methods in terms of quantitative metrics, visual quality and user study DiOr which supports 2D pose transfer, virtual try-on, and several fashion editing tasks. The key to DiOr is a novel recurrent generation pipeline to sequentially put garments on a person, so that trying on the same garments in different orders will result in different looks

FUTURE SCOPE

There many shopping websites which are adopting this technology and implementing this technology in software and even top MNC like lenskart adopt this technology and provide the facility of trying various products without going to the store .which make easy for customer to try any type of product so in future this technology is highly recommended.

REFERENCES

[1] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-EnWei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. IEEE transactions on pattern analysis and machine intelligence, 43(1):172–186, 2019.

[2] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014.

[3] Haoye Dong, Xiaodan Liang, Xiaohui Shen, Bochao Wang, Hanjiang Lai, Jia Zhu, Zhiting Hu, and Jian Yin. Towards multi-pose guided virtual try-on network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9026–9035, 2019.

[4] Haoye Dong, Xiaodan Liang, Yixuan Zhang, Xujie Zhang, Xiaohui Shen, Zhenyu Xie, Bowen Wu, and Jian Yin. Fashion editing with adversarial parsing learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8120–8128, 2020.

[5] Patrick Esser, Ekaterina Sutter, and Bjorn Ommer. A varia- ¨ tional u-net for conditional appearance and shape generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8857–8866, 2018

[6] Jean Duchon. Splines minimizing rotation-invariant seminorms in sobolev spaces. In Constructive theory of functions of several variables, pages 85–100. Springer, 1977.

[7] Jun Ehara and Hideo Saito. Texture overlay for virtual clothing based on PCA of silhouettes. In ISMAR, pages 139–142. IEEE Computer Society, 2006.

[8] Yuying Ge, Ruimao Zhang, Xiaogang Wang, Xiaoou Tang, and Ping Luo. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5337– 5345

A Review on Virtual Try on

Tripti Agrawal[1], Shitanshu Goyal[2], Shivam Patil[3]
Poornima College of Engineering,302022, India
2020pcecstripti187@poornima.org,
2020pcecsshitanshu173@poornima.org,
2020pcecsshivam176@poornima.org

Mrs. Barkha Narang
Assistant Professor
Poornima College of Engineering, 302022, India
barkha.narang@poornima.org

Abstract—
Now days the popularity of online shopping is increasing day by day Despite that growth stores have retained their main advantage – customers being able to try out products before buying them. But as shopping increasingly moves online, brands are striving to offer a holistic user experience on their digital channels too. Luckily, the advancing virtual try-on technology makes it possible to try on almost anything in just a few seconds – from makeup, jewellery and glasses to clothes and shoes. Prior arts usually focus on preserving the character of a clothing image (e.g. texture, logo, embroidery) when warping it to arbitrary human pose. However, it remains a big challenge to generate photo-realistic Tryon images when large occlusions and human poses are presented in the reference person. To address this issue, this work was done when Han Yang was a Research Intern at Sense Time Research. Adaptive Content Generating and Preserving Network (ACGPN) a visual try on network which first predicts semantic layout of the reference image which will be changed after try-on (e.g. long sleeve shirt→arm, arm→jacket), and then image content needs to be either generated or preserved will be determined By using the predicted semantic layout, it leads to photo-realistic try-on and rich clothing details. ACGPN generally involves three major modules.

I. Introduction

Nowadays electronic devices like smartphones, computers support Augmented Reality and the popularity keeps growing every day, mostly thanks to social networks. This makes virtual try-on solutions easily accessible and highly desirable to end users. It’s a big opportunity for all the companies or brands on favourite platforms to get closer to the customers. To stimulate the visualisation of dressing the demand of developing virtual dressing rooms is increasing.

Therefore, most researchers in previous works are taking the approach to map a 2D texture to the user’s body, and build an Avatar (model). In this work, we introduce a virtual dressing room application using Microsoft Kinect sensor. As the user stands in the front of the Kinect, his sizes measuring in real time, image mapping occurs. Further in this paper we have introduced VDRS i.e. Virtual Dressing Room system (VDRS), the main objective is develop an application that realistically reflects the look and feel of the clothes as it is supposed to. The clothes should adapt to certain bodies of different people.

It tough task to build up the photo-realistic virtual try-on system to apply in real world as to overcome such limitations Adaptive Content Generation and Preservation Network (ACGPN) is introduced which first predicts the semantic layout of the reference image and then adaptively determines the content generation or preservation according to the predicted semantic layout. The system incorporates a novel recurrent generation module to produce different looks depending on the order of putting on garments because of that it is named as DiOr, for Dressing in Order. In this research, we offer LA-VITON, an image-based virtual try-on network that outperforms earlier approaches in terms of providing pleasing-looking outputs without damage or distortion. Because GAN-based approaches have limits in terms of creating precise and clear pictures, an operation like GMM must be used to transmit data straight from the source to the target. The try-on photographs are smoothly and convincingly blended with the provided in-store apparel images and target human images.

1.1 Microsoft Kinect Sensor -

The application of virtual dressing room is using Microsoft Kinect sensor's as proposed approach is mainly based on extraction of the user from the video stream, alignment of models and skin colour detection. Modules are being used for locations of the joints for positioning, scaling and rotation in order to align the 2D cloth models with the user. Then, skin colour detection on video to handle the unwanted occlusions of the user and the model was applied. At the end the model is superimposed person who is using in real time. The problem is simply the alignment of the user and the cloth models with accurate position, scale, rotation and ordering. First, detection of the user and the body parts is one of the main steps of the problem. In literature, several approaches are proposed for body part detection, skeletal tracking and posture estimation, and superimposing it onto a virtual environment in the user interface. Kinect driver's middleware are used for the tracking process in combination with Microsoft and for various fundamental functions Kinect. As importance of Microsoft Kinect image sensor in the market is increasing so for that it is being applied and WFP to capture the user physical measurements which are as follows: -

1.1.1 *Kinect General components*: The components of Kinect for Windows are mainly the following:

- Kinect hardware: including the Kinect sensor
- USB hub, through which the sensor is connected to the computer
- Microsoft Kinect drivers: Windows 8 drivers for the Kinect sensor

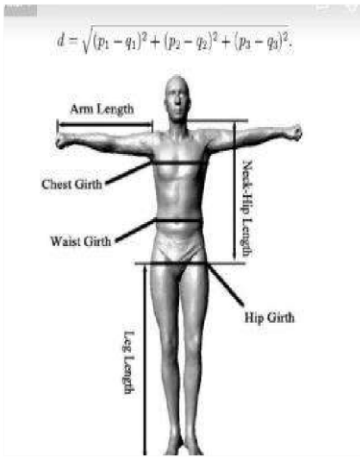
1.1.2 *WPF Application*: WPF stands for Windows Presentation Foundation which is UI framework to create applications with a rich user experience. It is part of the .NET framework 3.0 and higher. Its vector-based rendering engine uses hardware acceleration of modern graphic cards. This Makes the UI faster, scalable and resolution independent. In this section we will introduce the windows presentation foundation and discover its components and the features of each component as well.

In addition, the application keeps calculating distances to continue body movement:

The key points are as follows: -

1. Distances between joints positions

Using a general distance formula in 3D, we formulate an algorithm to all body distances. Each user enters the application in real time.



2. Convert Data:

Units of distance calculated using joint -joint coordinates results in metres so we have to convert it to pixel Data.

Using m – pixel converter

1 m = 3779.527559055 pixels.

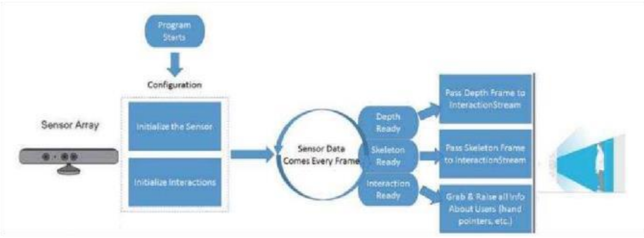
3. Create new joint positions:

Sometimes, key points that the Sensor found it, we face many problems in hip center, hip left, hip right points and it imprecise and inappropriate to what we need, in order to resolve this issue, we decided that we could make new points away from the original points left or right or down.

4. 2D cloth: We divide the clothes into parts of pixels data to control the movement of the body and the cloth in upper, lower frame. Using Photoshop CS6: we cut, divide, coloured the cloth to simulate it to a real body.

5. Interaction between human and 2D cloth:
• First, we take skeleton Data.
• Second, we take Depth Data.
• Third, we take the RGB data
• Fourth, we measure the upper Frame and mapping the
2D cloth.
• Fifth, we measure the Lower Frame and mapping the
2D cloth.
• Sixth, we measure the Lower, upper Frame and mapping the 2D cloth for both.

1.1.3 *Microsoft Kinect SDK version 1.0*: This version supports up to four Kinect sensors on a single computer, skeletal tracking, a Near Mode feature can make camera able to recognize objects just 40cm away, stability and audio, Drivers and runtime. The following diagram shows the components of the Kinect for Windows developers Toolkit.



1.2 ADAPTIVE CONTENT GENERATING AND PRESERVING NETWORK(ACGPN)

Adaptive Content Generating and Preserving Network which is used to preserve the character of clothes as well as the posture of person as to show the précised image. ACGPN generally involves three major modules.

1. A semantic layout generation
2. Clothes warping module
3. For content fusion module named as inpainting

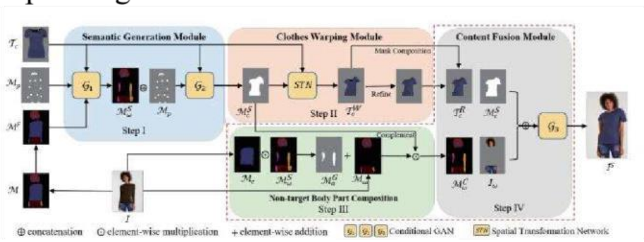


Fig: The overall process of ACGPN involving its three models

1.2.1 *Semantic Generation Module (SGM)*;- It is basically constructs the mask of various body parts as well as the clothes which are to be wrapped.

1.2.2 *Clothes Warping Module (CWM)*:- This module is made to warp the targeted clothing image according to the mask generated for wrapping the clothes and author also introduce an updated version or a second-order difference constraint on Thin-Plate Spline (TPS) to produce a geometric matching yet preserving the character of clothing images.

1.2.3 *Content Fusion Module (CFM)*:- If go further in deep or beyond the semantic alignment and character retention , it remains challenging task to realize layout adaptation on the task. Both the target clothing region is mandatory to get clearly rendered, and fine-scale details of various body parts (for example finger gaps) are needed to be adaptively Target Clothes

To preserve the character of clothes as well as the posture of person as to get the fine details of the clear and précised image adaptive content generating and preserving Network is introduced, i.e., it consists three carefully designed modules, i.e., Mask Generation Module, Clothes Warping Module and Content Fusion Module (CFM). The proposed approach is being evaluate their ACGPN on the VITON dataset by categorising them into three difficulty levels. The results proved that the ACGPN dominates or we can say had a great superiority over the state-of-the-art methods in terms of quantitative metrics, visual quality and user study.

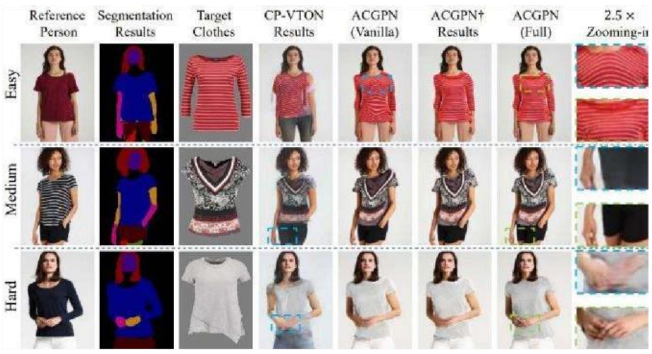


Fig: Categorising the try on task into three difficulty levels

By using ACGPN it greatly changed the quality as it improves a lot in various segments such as semantic alignment, character retention and layout adaptation. The higher order difference increases the stability of the training process of wrapping module and it can handle complex textures on clothes as it increases the ability of the model to handle such complex situations.

ACGPN, the current existing try-on methods mainly concentrate on keeping the posture, identity and character of clothes. There are some methods such as VITON, CP-VTON which by using coarse human shape and pose map as an input to generate a image of a person who is wrapped with clothes in precise manner. Whereas some methods such as SwapGAN, SwapNet and VTNFP adopt semantic segmentation as input to synthesize clothed person.

The results proves that the ACGPN show great advantage or more beneficial over other methods in terms of quantitative metrics, visual quality and user study. ACGPN as compared to the other methods can generate photo-realistic images with much better perceptual quality and richer fine-details.

1.3 DRESSING IN ORDER(DiOr)

Dressing in Order (DiOr) is a flexible person generation framework, which supports 2D pose transfer, virtual try-on, and several fashion editing tasks. The benefit or advantage of DiOr over other methods is that garments can be put on a person in a sequence so that if person try same garments in different order or poses it will result in different looks. DiOr very finely detects the shape and texture of each garment, so that it enables various elements to change separately. Joint training on pose transfer and inpainting helps with detail preservation and coherence of generated garments. On evaluating it extensively it clearly shows that DiOr shows great superiority over other methods which are came recently like ADGAN in terms of output quality, and handles a wide range of editing functions for which there is no direct supervision

DiOr is a flexible person generation pipeline trained on pose transfer and inpainting but capable to layer a wide range or diverse garments and the tasks for which there is no direct supervisions also edited. In general, the shading, garment detail preservation and texture warping of our method, are far better than those of other recent methods and still it is not realistic.

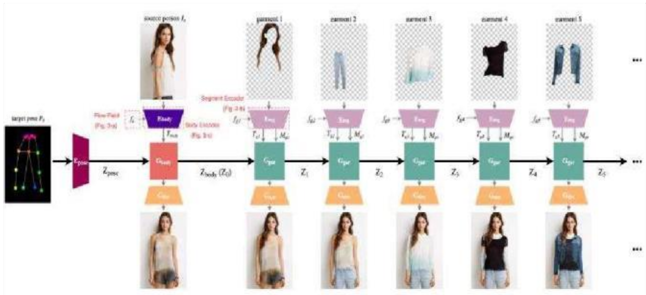


Fig: DiOr Generation pipeline

1.3.1 Generation Pipeline

In the main generation pipeline, firstly the “skeleton” P is to be encoded and after then generating the body from T 0 body and then the garments from texture which is encoded previously and shape masks (T 0 g1 , Mg1), ...,(T 0 gK , MgK) are to be taken in sequence. Pose and skin generation. To start generation, we encode the desired pose P by the help of pose encoder Epose, is to be implemented as three convolutional layers, each followed by instance normalization and leaky ReLU . This results in hidden pose map Zpose ∈ R L×H/4×W/4 , where L denotes the latent channel size and after that we make or generate the hidden body map Zbody given Zpose and the body texture map T 0 body by a body generator . by using two style blocks in ADGAN we can implement Gbody . Because our body texture map T 0 body is in 2D, SPADE. Replace the ADGAN’s adaptive instance normalization in the style block. DiOr was very much inspired by ADGAN. Same as ADGAN, we each garment can be encoded separately, condition the generation on 2D pose, and train on pose transfer. ADGAN encodes a garment into single 1D vector, but in terms of texture and shape garment is encoded separately in 2D. DiOr allows shape and texture of individual garments to be

Fig: Applications supported by DiOr system

edited separately where as in ADGAN it is not possible always filled in properly. For capturing complex spatial patterns Our 2D encoding is better than ADGAN's 1D encoding which giving us superior results on virtual try-on, as shown in next section. Besides, In ADGAN, after garments are separately encoded, all the embeddings are fused into a single vector, so the number and type of garments are fixed, and garment order is not preserved. By contrast, in our recurrent pipeline, garments are injected one at a time, and 14641 their number, type, and ordering can vary.

1.4 GEOMETRIC MATCHING MODULE

We use the bottom-up technique suggested by Rocco et al. for implementing the end-to-end trainable geometric matching (GMM). To improve GMM's performance, a matching grid constraint, an occlusion management strategy, and GAN loss are used.

1.4.1 *Grid Interval Consistency* - TPS transformations have a good track record, but their high flexibility can lead to pattern and print distortion. TPS with a high degree of freedom causes undesirable warping even when the coarse-to-fine strategy is used. We introduce grid interval consistency (GIC loss), which helps clothing preserve their qualities after warping.

1.4.2 *Occlusion Handling* - Hair and arms readily obscure a person's attire in a real-life setting. To solve the problem, we remove occlusion zones from the Lwarp computation. This technique helps the network to be taught to estimate transforming parameters with greater precision and reasonableness.

1.4.3 *Try on Module*

For the input person-image TOM synthesizes the final try-on image. To preserve the original details of in-shop clothing items, warped clothing should be exploited as much as possible. By blending the warped clothing from GMM and the intermediate person with a composition mask a seamless image is obtained. By following the generator, refinement layers improve the quality of the blended image higher. Dilated convolutions are used in the refinement layers to maintain the high-resolution feature map and to preserve picture details. for increasing the quality of pictures created using try on module SNGAN is used for training.

1.4.4 *Experiments*

Han et al. obtained a paired dataset, which we use. There are 14,221 training and 2,032 testing pairings in this dataset. The suggested GMM's alignment results are compared to those of the prior alignment methods. The suggested solution preserves pattern and print properties while also precisely aligning in-store clothing. For the first time, LA-VITON was compared to VITON and CP-VTON to evaluate if in-store garments had the same patterns and designs as those purchased in stores.

II. DISCUSSION

The results prove the various advantages of ACGPN, DiOr, Kinect sensor, VDRS over the state-of-the-art methods in terms of quantitative metrics, visual quality and user study. In comparison to the methods, ACGPN can generate photo-realistic images with clear and much better quality and richer fine-details. DiOr is a flexible person generation pipeline trained on pose transfer and inpainting but capable of diverse garment layering and editing tasks for which there is no direct supervision those of other recent methods, are still not entirely realistic. In the future, we plan to work on increasing the quality of the image produces by wrapping the clothes over the person through more advanced warping and higher-resolution training and generation.

III. CONCLUSION

To preserve the character of clothes as well as the posture of person as to show the précised image Author propose a novel adaptive content generating and preserving Network, i.e. Author present three carefully designed modules, i.e. Mask Generation Module (GMM), Clothes Warping Module (CWM), and Content Fusion Module (CFM). ACGPN is evaluated on the VITON dataset by categorising them into three levels of try-on difficulty. The results proves that ACGPN has great superiority over the state-of-the-art methods in terms of quantitative metrics, visual quality and user study. We can expect to see many more innovative uses for both technology in the future and perhaps a fundamental way in which we communicate and work thanks to the possibilities

IV. REFRENCES

[1] Remi Brouet, Alla Sheffer, Laurence Boissieux, and Marie-´Paule Cani. Design preserving garment transfer. ACM Trans. Graph., 31(4):36:1–36:11, 2012.

[2] Szu-Ying Chen, Kin-Wa Tsoi, and Yung-Yu Chuang. Deep virtual try-on with clothes transforms. In ICS, volume 1013 of Communications in Computer and Information Science, pages 207–214. Springer, 2018.

[3] Yunjey Choi, Min-Je Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In CVPR, pages 8789–8797. IEEE Computer So-ciety, 2018.

[4] Haoye Dong, Xiaodan Liang, Bochao Wang, Hanjiang Lai, Jia Zhu, and Jian Yin. Towards multi-pose guided virtual try-on network. CoRR, abs/1902.11026, 2019.

[5] Haoye Dong, Xiaodan Liang, Yixuan Zhang, Xujie Zhang, Zhenyu Xie, Bowen Wu, Ziqi Zhang, Xiaohui Shen, and Jian Yin. Fashion editing with multi-scale attention normaliza-tion. CoRR, abs/1906.00884, 2019.

[6] V. Scholz and M. Magnor, “Multi-view video capture of garment motion,” in Proc. IEEE Workshop on ContentGeneration and Coding for 3D-Television, 2006, pp. 1–4.

[7] A. Del Bue and A. Bartoli, “Multiview 3d warps,” inComputer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 675–682.

[8] J. Shen, X. Yan, L. Chen, H. Sun, and X. Li, “Re-texturing by intrinsic video,” Information Sciences, vol. 281, pp. 726–735, 2014.

[9] A. Hilsmann and P. Eisert, “Optical flow based tracking and retexturing of garments,” in Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. IEEE, 2008,pp. 845–848.

[10] http://glamstorm.com/en/fittingroom/clothes# .

[11] http://www.metail.com.