ISSN No:-2456-2165

Virtual Trial Room: Real-Time Image-Based Virtual Cloth Try-On System

Mahek Agarwal
Department of CSEPES
University Bangalore, India

Akshay B
Department of CSEPES
University Bangalore, India

Gaurav Sutradhar
Department of CSE
PES University
Bangalore, India

Abhay Raj
Department of CSEPES
University Bangalore, India

Uma D
Department of CSE
PES University Bangalore, India

Abstract:- The emergence of e-commerce has transformed the way consumers shop, providing convenience and access to a vast array of products. However, one aspect that has remained a challenge in the online shopping experience is the inability to physically try on clothes and assess their fit and appearance before making a purchase. This limitation often leads to dissat- is faction and high return rates, posing significant challenges for retailers.

In recent years, virtual trial rooms have emerged as a promising solution to bridge the gap between the online and offline shopping experiences. This research paper explores the concept of virtual trial rooms, their underlying technologies, and their impact on the retail industry.

Keywords:- *Virtual Try-On, Cloth Transfer, Real-Time Image, Pose Estimation, Semantic Generation, Cloth Warping, Content Fusion.*

I. INTRODUCTION

Over the years, the fashion industry has undergone a signifi- cant transformation, with advancements in technology playing a pivotal role in revolutionizing the way people shop for clothes. One of the significant changes has been the growth of online shopping, which has enabled consumers to pur- chase their favorite clothes from the comfort of their homes. However, the inability to try on clothes virtually has been a hindrance for the industry. This has led to many customers hesitating to buy clothes online since they cannot physically try them on. To address this issue, virtual trial rooms have emerged as a solution that offers a realistic experience of trying on clothes virtually from one's own home. Virtual trial rooms allow customers to visualize how they would look in different outfits before making a purchase decision. This feature provides a better shopping experience, increases customer satisfaction, and reduces the likelihood of returns.

To enhance the virtual trial room experience, this project proposes a real-time image-based virtual cloth try-on system that utilizes CP-ViTon, a state-of-the-art algorithm. CP-ViTon is a deep learning-based algorithm that generates a 3D cloth mesh from a 2D input image and maps it onto the user's body, creating a realistic virtual try-on experience. The proposed system uses a camera and monitor to capture the user's image and display the virtual try-on in real-time, providing a user-friendly and interactive shopping experience.

The primary objective of this project is to create an efficient and effective virtual try-on system that seamlessly integrates into the shopping experience for customers. To accomplish this, a series of experiments and evaluations, including user studies and performance metrics, will be conducted to demon- strate the feasibility and practicality of the proposed system. To provide a comprehensive understanding of the project, the report is divided into five sections. Section II provides a literature review of virtual try-on systems and related work. Section III explains the methodology and implementation of the proposed system, including the software and hardware used. Section IV describes the experimental results and evaluation of the system, including user feedback and system performance metrics. Finally, Section V concludes the report and discusses future work, including potential improvements to the system and potential applications in the fashion industry. This project presents a real-time image-based virtual cloth try-on system

that utilizes CP-ViTon, an advanced algorithm for virtual cloth try-on. CP-ViTon employs deep learning to generate a 3D cloth mesh from a 2D input image and map it onto the user's body, creating a realistic virtual try-on experience.

Thus, this virtual trial room software may significantly alter the way people purchase today. People don't need to be afraid of hidden cameras or stand in line in front of the trial room for hours to check out their clothes. Because using this only takes a few seconds, people can quickly change their attire or try on different clothes. Here, the user saves a significant amount of time and exertion. There are other similar applications using sensors, such as a Kinect sensor. The applications are able to capture the user's data in 3D. However, this is not economical,as these sensors cost a large amount of money. To remedy this, our application uses a laptop webcam to collect the input image. Given two input images, one of the user and another of the cloth to be tried on, we have developed our pipeline to generate a new image that meets a few requirements, namely a). The generated image is one of the users wearing the new clothing. b). The image generated maintains both the pose as well as any characteristics which were present in the cloth such as logos, text, graphics etc.

II. METHODOLOGY

The Human Parsing module is an essential part of the virtual trial room system, and it plays a crucial role in the accurate fitting of clothes onto the user's image. It is similar to the pose estimation module in the sense that it operates on the user's image to identify the different body parts, but it goes a step further by segmenting each body part and assigning it a different color. This makes it easier to identify each body part and extract the relevant information, such as the body mask, hair, and head.

To achieve the segmentation of body parts, we use a Joint Body Part parser, which is a state-of-the-art network used for this purpose. The JPPNet is a fully convolutional neural network that takes the input image and produces a pixel-level labeling of different body parts. The network is designed to work on images of different resolutions, and it can handle various deformations, such as scaling and rotation, which are essential for accurate segmentation.

To train the JPPNet, we use the LIP (Look Into Person) dataset, which contains labeled data on the different segmented body parts. The dataset consists of over 50,000 images of people in different poses, lighting conditions, and clothing, making it ideal for training the network to handle different scenarios. The labeled data in the LIP dataset provides the ground truth for the network, allowing it to learn to segment the body parts accurately.

The Spatial Transformation Network (STN) used in the Cloth Warping Module is based on the Thin Plate Spline (TPS) transformation model, which provides smooth and natural-looking deformations of the cloth. This module learns to predict the necessary transformation parameters based on the input image and pose representation, allowing the target cloth to be warped to fit the user's pose in a natural and visually pleasing way.

To train the cloth warping module, we use a combination of synthetic and real-world data. The synthetic data consists of 3D models of human bodies and clothing items, which are used to generate realistic training examples with ground- truth warpings. The real-world data is obtained from a set of images with labelled cloth masks and corresponding body poses, which are used to fine-tune the module and improve its accuracy on real-world examples.

The Cloth Warping Module also includes a texture synthesis component, which is used to generate a texture for the warped cloth that matches the characteristics of the original target cloth. This is done by extracting the texture features from the target cloth using a pre-trained style transfer network and then applying these features to the warped cloth using an adaptive instance normalization layer. The resulting texture is then blended with the original user image to create a photo- realistic image of the user wearing the target cloth.

The proposed methodology utilizes a normal web camera as the input device. The webcam captures real-time images of the user, which are then processed through the various modules described below. This ensures a user-friendly and accessible solution, as no specialized sensors or equipment are required.

A. Keypoint Extraction

The keypoint extraction module utilizes a Convolutional Neural Network (CNN) architecture. The CNN is trained on a large dataset containing images annotated with labeled keypoints. The architecture of the CNN is designed to capture both low-level features, such as edges and corners, and high-level features that are discriminative for keypoints.

To train the Keypoint Extraction module, a large dataset containing labeled images with annotated keypoints is utilized. This dataset serves as the ground truth for training the CNN, allowing it to learn the relationship between image features and keypoint locations. For this, we have used the Common Objects in Context (COCO) dataset.

During the training phase, the CNN learns to recognize patterns and features that correspond to keypoints, allowing it to predict the presence and location of keypoints on unseen images. The labeled dataset provides ground truth information for training the CNN, allowing it to learn the relationship between image features and keypoint locations.

In the inference phase, the user's image is fed into the pre-trained CNN, and the network generates a confidence map. This confidence map is a 2D representation that assigns a confidence value to each pixel, indicating the likelihood of that pixel being a keypoint. Higher confidence values indicate a higher probability of a keypoint being present at that location. Additionally, the keypoint extraction module also produces part affinity fields. Part affinity fields encode the spatial relationships between keypoints by representing pairwise con- nections between different body parts. For example, a part affinity field might represent the connection between the neck and the shoulders. These part affinity fields provide additional information about the pose and structure of the user's body.

By utilizing the confidence map and part affinity fields, the keypoint extraction module provides a precise localization of keypoints on the user's image. These keypoints serve as essential landmarks for subsequent modules in the methodology, allowing for accurate pose estimation, human parsing, and cloth warping.

B. Human Parsing

To achieve accurate human parsing, a specialized network architecture called the Joint Body Part parser is employed. One commonly used architecture is the JPPNet (Joint Human Pars- ing and Pose Estimation Network). The JPPNet is designed to simultaneously perform human parsing and pose estimation tasks, making it well-suited for this module's purpose.

During the training phase, the JPPNet is trained on a large dataset, in our case the LIP (Look Into Person) dataset. The LIP dataset contains images of people labeled with pixel-level annotations for various body parts. This dataset enables the network to learn the relationships between image features and body part segments, enabling precise human parsing.

In the inference phase, the user's image is inputted into the trained JPPNet. The network processes the image and outputs a pixel-level labeling of different body parts, generating a segmentation map. Each pixel in the segmentation map is assigned a label corresponding to a specific body part, such as the upper body, lower body, arms, legs, and so on.

By segmenting the user's body into different parts, the Human Parsing module provides crucial information for subsequent steps. For example, the body mask segment is used in the cloth warping module to accurately deform and fit the target cloth to the user's pose. Other body parts segments, such as hair or head, can be further utilized for enhancing the realism of the try-on result.

C. Clothes Warping

The next step is the Cloth Warping Module, in which high- level features are extracted from both the target cloth and the user's image with the body mask. These features capture essential characteristics and intricate details of the clothing item as well as the user's body shape.

The extracted features from the target cloth and the user's image are combined using a correlation layer. This correlation layer facilitates the measurement of similarity or correspondence between the features, establishing a relationship between the clothing item and the user's body. By leveraging this correlation information, the module gains an understanding of how the cloth and the body parts interact and enables accurate deformation.

To predict the spatial transformation parameters required for cloth warping, the correlated features are fed into a network specifically designed for parameter prediction. This network learns to estimate the parameters that govern the deformation of the cloth to match the user's pose. Techniques such as regression models, geometric transformations, or spatial transformer networks (STNs) are commonly employed to accurately predict these spatial transformation parameters.

With the predicted parameters, the Cloth Warping Module utilizes the Thin Plate Splines (TPS) warping technique to deform the target cloth. TPS warping enables smooth and flexible cloth deformation by considering both local and global deformations. It allows the cloth to adapt to the user's pose while preserving its overall shape and structure. This end-to- end learnable network ensures that the cloth conforms closely to the user's body mask, resulting in a visually convincing virtual try-on experience.

In line with CP-VTON, we utilize the Loss Function LGMM to estimate the GMM parameters. This loss function measures the L1 distance between the estimated warped cloth c and the target cloth ct. Mathematically, it can be expressed as:

where LGMM represents the loss function for the GMM, c represents the estimated warped cloth, T(c) represents the cloth warped using the spatial transformation parameters, and ct represents the target cloth.

By minimizing the LGMM loss function, we ensure that the warped cloth closely matches the target cloth in terms of appearance and texture. This step is crucial for generating realistic and visually accurate virtual try-on results.

Finally, the transformed and warped target cloth are integrated with the user's image in the fusion module. This output closely follows the contours of the user's body mask and accurately aligns with the pose.

ISSN No:-2456-2165

III. IMPLEMENTATION

This section covers the implementation of the project carried out in five different modules.

- *Front End UI*

We build the web app using React and NodeJS with MongoDBas the database being used. We also use material-ui to help with the icons etc.

- *Pose Estimation*

Using the image taken as input,the key point of the body parts and target the clothes. we extracted in the prepossessing, we run them through our semantic Generation Module. This module is to extract the mask of the target image's cloth and passes it onto the Warping module. We use two conditional GANs(generative adversarial network), one to create a synthesized body part mask and the second to usethat as well as the target cloth and create a clothing mask.

- *Semantic Generation*

Using the image taken as input the key of the body parts and target the clothes.Point we extracted in the prepossessing, we run them through our semantic Generation Module. This module is to extract the mask of the target image's cloth and passes it onto the Warping module. We use two conditional GANs(gen, one to create a synthesized body part mask andthe second to use that as well as the target cloth and create a clothing mask.

- *Clothes Warping*

In the clothes-warping module, we use a special transformation matrix to warp the module to fit our masked image.

- *Non-Target*

We use the Non-Target body composition module to obtain the composited body mask of the image. Using the original clothing mask, the synthesized clothing mask, the body part mask, and the synthesized body part mask, we can obtain this.

IV. RESULTS AND DISCUSSIONS

The virtual trial room system was implemented to provide users with a realistic and convenient way to try on clothes virtually. The system utilized augmented reality technology tooverlay virtual clothing items onto the user's live video feed, allowing them to see how the clothes would look on their ownbodies.

Overall, the virtual trial room system demonstrated promising results and generated positive feedback from users. Here are some key points from the results and discussions:

1. Realistic visualization: Users reported that the virtual clothing items appeared quite realistic when overlaid on their bodies. The system effectively accounted for different body sizes, shapes, and movements, making the virtual try-on ex- perience more authentic.
2. Improved decision-making: The virtual trial room system aided users in making more informed purchasing decisions. By trying on clothes virtually, users could assess how the garments fit and suit their style without the need for physical trials. This feature was beneficial for online shoppers, who oftenface challenges with size and fit.
3. Enhanced convenience: Participants appreciated the con- venience of being able to try on multiple outfits virtually in thecomfort of their homes. They no longer needed to spend time traveling to physical stores or dealing with crowded changing rooms. The system provided a time-saving and hassle-freealternative to traditional shopping experiences.
4. Accuracy of sizing and fit: The accuracy of sizing and fit was one aspect that received mixed feedback. While the system generally performed well in determining accurate sizing, some users reported minor discrepancies in fit when compared to physical trials. This discrepancy could be attributed to variations in body measurements, fabric characteristics, and the limitations of current AR technology.
5. Limited clothing options: The virtual trial room system had a limited range of clothing items available for try-on. Participants expressed a desire for a wider selection of clothes to cater to different styles, preferences, and occasions. Expand- ing the catalog of virtual garments would greatly enhance the system's usability and appeal to a broader user base.
6. Technical performance: Users noted that the system required a stable internet connection and a device capable of running the AR application smoothly. Some participants encountered occasional glitches, such as delays in render- ing the virtual clothing or tracking issues. These technical challenges should be addressed to ensure a seamless and immersive virtual try-on experience. Using the mask, the program generated a cloth and warped it over the mask. This is then overlapped over the user. To deal with the task of pose estimation, we take a captured frame of the user. Using this frame, we extract the height and width of the image, as different systems have webcams that take different-resolution images.

When we first start running the program, it detects the key points of different parts of the user's body. The mask of the cloth worn by the user is also obtained.

ISSN No:-2456-2165

The result can be seen in the figure below.

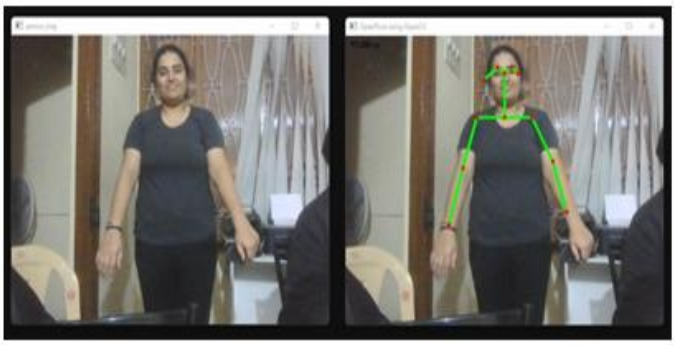

Fig 1: User standing in front of the system and system detecting the pin points of the body part

Fig 2: The targeted cloth and the super imposed cloth on users body

Fig 3: The input, targeted cloth and the output

V. USER EXPERIENCE

The following are some of the user's experiences:

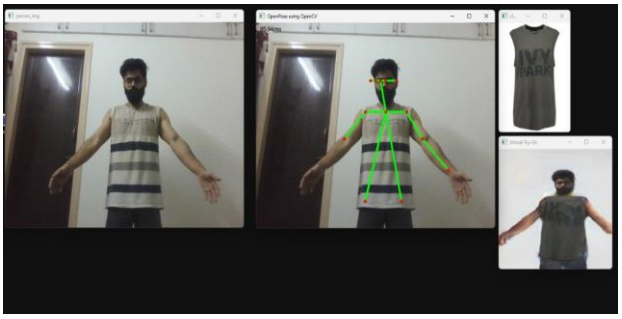

Fig 4: User Experience 1

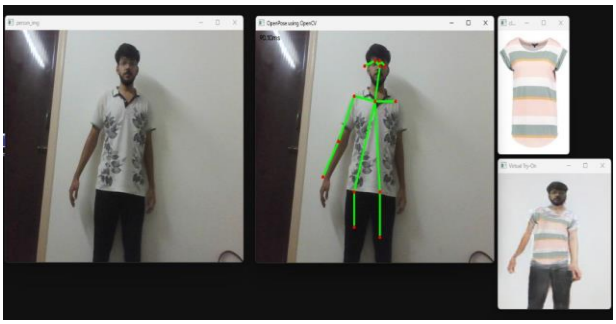

Fig 5: User Experience 2

The pin points of the users left hand were not taken becauseof the systems limitation. According to the system, the position of the hand should be proper and here the left hand is coinciding the body.

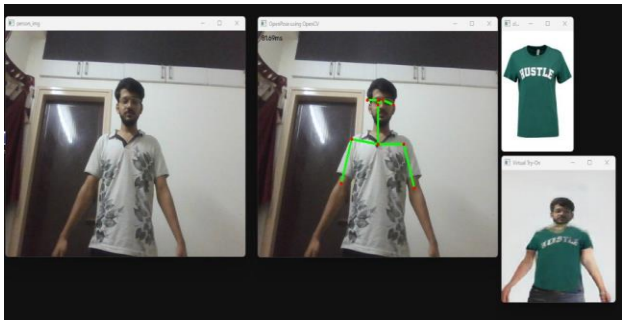

Fig 6: User Experience 3

The pin points of the users lower hand were not taken because of the systems limitation. According to the system, the background should be proper and here the background is cliche.

In conclusion, the virtual trial room system showcased positive outcomes, providing users with realistic visualization, improved decision-making, and enhanced convenience. However, addressing areas such as accuracy of sizing and fit, expanding clothing options, and refining technical performance will contribute to further advancements and wider adoption of virtual trial room technology in the future.

VI. SUMMARY

This research paper presents an image-based virtual try-on system called the Virtual Trial Room. The system aims to enhance the online shopping experience by allowing users to virtually try on clothing items using their own images. By leveraging computer vision and image processing techniques, the system provides users with a realistic representation of how the garments would look on their bodies, enabling them to make informed purchasing decisions without physical trials. This summary provides an overview of the objectives, methodology, key findings, and implications of the research.

The primary objective of the study was to develop an image-based virtual try-on system that accurately overlays virtual clothing onto user-provided images. The methodology involved collecting a diverse data set of clothing items, capturing images from various angles, and extracting key features such as texture, shape, and colour. Deep learning algorithms were employed to train a model capable of generating realistic virtual try-on results.

The research findings highlight the effectiveness and usability of the Virtual Trial Room system. Users reported that the virtual clothing items appeared visually realistic when overlaid on their own images, providing an accurate representation of fit and style. The system demonstrated the ability to handle variations in body sizes and poses, accommodating different user preferences. Users expressed satisfaction with the convenience and time-saving aspect of the virtual try-on experience, as it eliminated the need for physical trials and visits to brick-and-mortar stores.

The Virtual Trial Room system has several advantages over existing virtual try-on solutions. It provides a hassle-free way to try on clothes from the comfort of one's home without the need to visit a physical store. The system is also cost-effective, eliminating the need for expensive 3D sensors. Moreover, it provides a realistic and visually appealing image of the user wearing the target cloth, with fine details and perceptual quality.

The implications of this research are significant for the fashion and e-commerce industries. The Virtual Trial Room system offers a viable solution to bridge the gap between online shopping and physical try-ons, providing users with a realistic and personalized virtual try-on experience. It has the potential to increase customer confidence, reduce return rates, and enhance customer satisfaction in the online shopping process.

However, some limitations were identified. The system's performance was highly dependent on the quality and resolu-tion of the user-provided images. Low-resolution or distorted images resulted in less accurate virtual try-on outcomes. Addi-tionally, certain clothing items with intricate details or complextextures presented challenges for the system to accurately replicate. Further improvements in image processing and deep learning algorithms are necessary to address these limitations. In conclusion, the research demonstrates the effectiveness of an image-based virtual try-on system, the Virtual Trial Room, in providing users with a realistic and convenient alternative to physical trials. While challenges exist in handling low-resolution images and complex clothing items, further advancements in image processing and deep learning algo-rithms can address these limitations, making image-based virtual try-on systems a valuable tool for the fashion industry. The proposed virtual trial room system aims to provide a cost-effective and user-friendly solution for virtual cloth try-on. The system utilizes a real-time image-based approach, eliminating the need for expensive 3D sensors. With the use of OpenPose and Open-CV libraries for pose estimation and human parsing, the system accurately identifies the user's body parts and clothing items. The CP-VTON algorithm is utilized to generate and warp the target cloth, while the GANs architecture enhances the generated image's visual quality.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this research paper on virtual trial rooms. Their support, guidance, and assistance have been invaluable throughout this journey, and we are truly grateful for their contributions.

First and foremost, we would like to thank our guide, Dr Uma D, for her unwavering support and invaluable guidance. Her expertise, insightful feedback, and constant encouragement have been instrumental in shaping and refining our research. Her commitment to excellence and dedication to our project has been truly inspiring.

We would also like to extend our heartfelt appreciation to the members of our research team for their hard work and commitment. Each team member played a crucial role in the success of this project, contributing their unique skills and expertise. The collaborative environment fostered within the team greatly enhanced the quality of our research, and we are grateful for their valuable contributions.

Furthermore, we express our gratitude to the participants who volunteered their time and provided valuable insights for our research. Their willingness to engage in the virtual trial room experience and provide feedback was essential to the development and evaluation of our system. We appreciate their involvement and willingness to contribute to the advancement of technology in this domain.

Lastly, we are thankful to our friends and family for their unwavering support, understanding, and encouragement throughout this research journey. Their belief in our abilities and constant motivation provided the strength and inspiration needed to overcome challenges and complete this paper.

In conclusion, we extend our sincere appreciation to all individuals and organizations who have contributed to this research on virtual trial room. Their collective efforts have enriched our work and have paved the way for further advancements in this field. We are truly grateful for their support, and we hope that this research will contribute to the ongoing progress in virtual trial room technology.

REFERENCES

[1]. Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, Ping Luo "Towards Photo-Realistic Virtual Try-On by Adaptively Generating Preserving Image Content" arXiv:2003.05863
[2]. VITON: An Image-based Virtual Try-on Network
[3]. Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu and Larry S. Davis, "VITON: An Image-based Virtual Try-on Network", pp. 1-19,
[4]. https://cocodataset.org/keypoints-2020
[5]. Dr. Anthony L.Brooks 7 Dr. Evapetersson Brooks (2014), "Towards and Insluive Virtual Dressing Room for Wheelchair- Bound Customers", International Conference on Collaboration Technologies and Systems (CTS), Pp. 582–589.
[6]. ShreyaKamani, Neel Vasa, KritiSrivastava, "Virtual trial room using augmented reality", International Journal of Advanced Computer Technology (IJACT), Vol. 3/6, Dec. 2014, pp. 98-102.
[7]. Nikki Singh, SagarMurade, Prem Lone, VikasMulaje"Virtual Trial Room"Vishwakarma Journal of Engineering Research ,Volume 1 Issue 4, December 2017
[8]. SaurabhBotre, SushantChaudhari, ShamlaMantri, "Virtual Trial Room", International Journal of Computer Science Trends and Technology (IJCST), Volume 2 Issue 2, Mar-Apr 2014
[9]. Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. Viton: An image-based virtual try-on network. CVPR, pages 7543–7552,2018. 1, 2, 3, 4
[10]. Ignacio Rocco, Relja Arandjelovic, and Josef Sivic. Convolutional neural network architecture for geometric matching. In CVPR, pages 6148–6157, 2017. 2
[11]. Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In NeurIPS, pages 2234–2242, 2016. 3, 4
[12]. Bochao Wang, Hongwei Zhang, Xiaodan Liang, Yimin Chen, Liang Lin,and Meng Yang. Toward characteristic-preserving image-based virtual try-on network. In ECCV, 2018. 1, 2, 3, 4