International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6942
***Article Received**: 26 March 2023 **Revised**: 18 May 2023 **Accepted**: 28 May 2023*

A Survey of Algorithms Involved in the Conversion of 2-D Images to 3-D Model

[1]Gopal D. Upadhye, [2]Anant Kaulage, [3]Ranjeetsingh S. Suryawanshi, [4]Roopal Tatiwar, [5]Rishikesh Unawane, [6]Aditya Taware, [7]Aryaman Todkar

Department of Computer Engineering,
Vishwakarma Institute of Technology, Pune
Pune, India
[1]gopal.upadhye@vit.edu
[2]anant.kaulage@vit.edu
[3]ranjeetsingh.suryawanshi@vit.edu
[4]roopal.tatiwar20@vit.edu
[5]rishikesh.unawane20@vit.edu
[6]aditya.taware20@vit.edu
[7]aryaman.todkar20@vit.edu

Abstract—Since the advent of machine learning, deep neural networks, and computer graphics, the field of 2D image to 3D model conversion has made tremendous strides. As a result, many algorithms and methods for converting 2D to 3D images have been developed, including SFM, SFS, MVS, and PIFu. Several strategies have been compared, and it was found that each has pros and cons that make it appropriate for particular applications. For instance, SFM is useful for creating realistic 3D models from a collection of pictures, whereas SFS is best for doing so from a single image. While PIFu can create extremely detailed 3D models of human figures from a single image, MVS can manage complicated situations with varied lighting and texture. The method chosen to convert 2D images to 3D ultimately depends on the demands of the application.

Keywords-2-D image, 3-D model, SFM, SFS, MVS, PIFu, Machine learning.

I. Introduction

A fascinating and quickly developing area of computer vision and image processing is 2D to 3D picture conversion. It entails taking a 2D image and turning it into a 3D representation, enabling better visualization and depth perception. The technology underlying this conversion process has a wide range of uses, from the entertainment sector to virtual reality and medical imaging. The quality and accuracy of 2D-to-3D image conversion have substantially increased thanks to recent developments in machine learning, deep neural networks, and computer graphics, making it a viable area for continued research and development.

For product design, prototyping, testing, and manufacturing, various sectors use 3D modeling. In order to visualize and plan building plans, architects and construction experts utilize 3D models. This enables them to find design defects and maximize efficiency prior to the start of construction. 3D models are also used in the automotive sector to develop car systems and parts, simulate vehicle performance, and assess crashworthiness. In the medical field, 3D models are used to simulate the activity of organs and tissues, plan and visualize procedures, and design unique implants and prosthetics. In addition, the entertainment sector makes extensive use of 3D modeling to produce realistic visual effects, animations, and video games. As 3D modeling technology continues to evolve, more businesses are utilizing it to streamline their processes, cut costs, and improve their goods and services.

The convenience and ease-of-use of currently available 3D-capable technology, such as Televisions, Blu-ray players, gaming consoles, and Smartphone’s, has not yet been matched by the development of 3D content. In contrast to 2D content, which has been around for decades and is widely available today, 3D content is rather hard to get by. This is mostly because creating 3D models is more difficult and time-consuming than creating 2D photographs, which can be done quickly using a camera or a digital drawing tool. The equipment and abilities needed for 3D modeling and rendering are also more difficult to obtain than those for creating 2D material. Due to the dearth of 3D content, many applications that would benefit from it, such virtual reality and augmented reality, are constrained. As a result, the capacity to transform 2D photos to 3D might be a game-changer by enabling the generation of new 3D material utilizing 2D content that already exists and so addressing the lack of 3D content.

This paper presents a systematic comparative study on current methods used to convert 2-D images to 3-D models. The techniques covered in the paper include:

- PIFu (Pixel-Aligned Implicit Function)

DOI: https://doi.org/10.17762/ijritcc.v11i6s.6942
***Article Received**: 26 March 2023 **Revised**: 18 May 2023 **Accepted**: 28 May 2023*

- PIFuHD (Pixel-Aligned Implicit Function HD)
- SFS (Shape from Shading)
- SFM (Structure from Motion)
- MVS (Multi-View Stereo)

II. Literature Review

For the creation of this paper, a lot of research was done. A few notable papers have been discussed in this section of the paper.

Natsume R et. al [1] presented a new technique Pixel-aligned Implicit Function (PIFu). It is an implicit representation that synchronizes the local context of a 2D image's pixels with that of its corresponding 3D object. They suggested a fully-featured deep learning technique that can derive texture and 3D surface from a single image and, optionally, from a set of input photographs for digitizing persons who are clothed elaborately.

After PIFu, an updated version was released in [2] by ChanE R. et. al. It was named PIFu hd. The difference between both the algorithms is that compared to PIFu, PIFu HD creates 3D reconstructions of higher quality and with more precise details. PIFu and PIFuhd are comparatively latest techniques that came into market around 2019-2020. But before these were introduced other techniques were used(like SFS, SFM, etc), which are also discussed in detail throughout the paper.

Xiang Wang et. Al. [10], presented a paper on multi-view stereo technique. The study detailed various widely used datasets and the accompanying metrics for evaluation, and it showed how MVS was implemented. Finally, a number of perceptive conclusions and difficulties are advanced illuminating prospective study options.

Kholil, Moch et. Al.[8], implemented 2 algorithms: MVS and SFM. In an effort to preserve three-dimensional artefacts in the Penataran Temple cultural heritage zone, they used these methods of three-dimensional model reconstruction. Up to 61 photos of various items in the vicinity of the BlitarPenataran Temple were taken for this study. The captured images were converted into a 3D model utilizing the meshroom's Structure From Motion (SFM)technique.

These are only a few of the many papers surveyed. A certain methodology was followed while doing the survey of papers. It can be observed in the diagram shown in Figure 1.

The survey started with a clear definition of problem statement. It was vital for the research, for us to understand the problem statement purpose and of this survey. After its careful study, papers related to the problem domain were researched. The main sources of our papers were – google scholar and arxiv. Both are digital libraries or archives that have countless research papers that researchers can refer to. The papers surveyed gave us an insight on the various methodologies and approaches that were implemented. We carefully studied these papers and listed down the observations and results. This can be observed in Table 1.

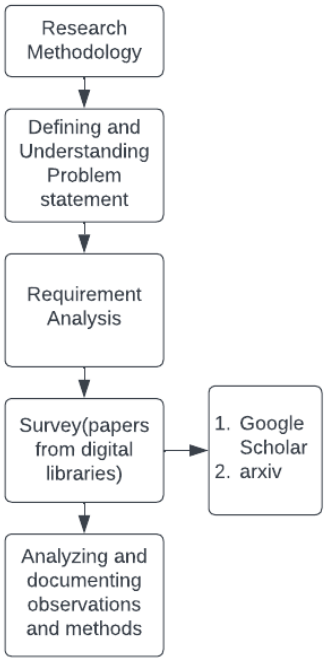
Research Methodology
Defining and Understanding Problem statement
Requirement Analysis
Survey(papers from digital libraries)
1. Google Scholar
2. arxiv
Analyzing and documenting observations and methods

Figure 1. Block Diagram for methodology implemented for survey.

III. METHODS

This section of the paper represents a brief overview of all the techniques that are used to implement the conversion of 2-D image to 3-D modeling.

A. PIFu

The deep learning system known as PIFu, or Pixel-Aligned Implicit Function, permits the reconstruction of a 3D representation of an object from just a single 2D photograph. The technique employs an implicit function to match the object's 3D geometry with the pixels of the image taken as input in order to produce the best results. The genesis of PIFu was described in [1].

PIFu makes use of a neural network architecture to produce an implicit function that depicts the object's 3D surface from a 2D image as input. As a continuous function that can be evaluated at any location in space, the implicit function serves as a representation of the object's three-dimensional surface. PIFu can produce a 3D model of the item that precisely replicates the appearance of the 2D input image by lining up the implicit function with the pixels of the input image. PIFu excels at creating 3D models of human subjects from a single photograph, whether the subjects are dressed or not. The algorithm can create high-resolution 3D models with precise details, including folds and wrinkles, and it can manage a variety of garment styles and positions. PIFu has many uses in

***Article Received**: 26 March 2023 **Revised**: 18 May 2023 **Accepted**: 28 May 2023*

a variety of industries, including video games, fashion design, and virtual and augmented reality. The PIFuHD algorithm is an improvement of the PIFu algorithm, which was first introduced in a research article by the same authors in 2020 and is also known as Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization [3]. PIFuHD was created primarily to enhance the reconstruction of highly detailed 3D models of clothed human people.

The ground truth surface for surface reconstruction is represented in PIFu as a continuous 3D occupancy field with 0.5 levels:

$$f_v^*(x) = \begin{cases} 1, & if\ X\ is\ inside\ mesh\ s\imath \\ 0, & otherwise \end{cases} \quad (1)$$

To create a pixel-aligned implicit function (PIFu) fv, the mean squared error is typically reduced:

$$L_v = \frac{1}{n}\sum_{i=1}^{n} |f_v(F_v(x_i), z(X_i)) - f_v^* \quad (2)$$

B. PIFuHD

Similar to PIFu, PIFuHD reconstructs 3D models from 2D pictures using a neural network architecture. A multi-level hierarchical structure that better captures the details of the 3D model, a novel texture mapping algorithm for synthesizing high-quality textures for the reconstructed model[51], and a feature that enables the reconstruction of multiple human subjects in a single image are just a few of the significant advancements made by PIFuHD.

The latter's capacity to handle photos with greater resolutions is one of the main distinctions between PIFu and PIFuHD. While PIFu is limited to 512 x 512 pixels, PIFuHD can produce 3D models with a resolution of up to 2048 × 2048 pixels. There are several uses for PIFuHD's enhanced capabilities, notably in the fashion and entertainment sectors. Also, PIFuHD's developments have helped the field of computer vision by making it possible to create 3D reconstruction algorithms from 2D images that are more precise and effective.

C. SFS

A traditional computer vision technique called SFS (Shape from Shading) determines an object's 3D shape based on the shading information in a 2D image. It presumes that the surface of the object is Lambertian, reflecting light uniformly in all directions. The fundamental tenet of SFS is that one may estimate the orientation of the surface normal's of an object at each pixel by using the shading information in an image. SFS can reconstruct the object's 3D geometry up to a scale factor by integrating these normal vectors.

The procedure usually assumes that the picture is illuminated by a single point light source and that the object has a smooth surface devoid of any sharp edges or occlusions in order to execute SFS. After making these presumptions, the algorithm analyzes the brightness gradients in the image to estimate the surface normal vector at each pixel. There are several SFS variations, including the Photometric Stereo technique, which determines the surface normal's using multiple images of the same object taken in various lighting conditions, and the Shape from Polarization technique, which determines the surface normal's using information about the polarization of the light. SFS is a useful method for creating a 3D model of an object from a 2D image, but it has its limitations. For instance, it has limitations when dealing with objects that have specular reflections or shadows and is sensitive to the assumption of a Lambertian surface. Because it is noise-sensitive, the parameters of the camera and light source must be carefully calibrated.

D. SFM

SFM (Structure from Motion), a popular photogrammetry technique, reconstructs the 3D structure of an item or scene by looking at a set of 2D images taken at various angles. The method determines the 3D coordinates of points on an object's surface by analyzing the relationships between points in several photos. SFM starts by determining the camera poses for each input image or the location/position and orientation of the camera with respect to the object. Usually, direct methods or feature matching are used to do this. Triangulation is then employed by the program to estimate the 3D coordinates of points on the object's surface by establishing the link between points in several pictures. Sparse reconstruction is one of the techniques used to implement the SFM approach.

With a collection of 2D picture observations, sparse reconstruction in SFM entails estimating the camera postures and sparse 3D point placements. In SFM, sparse reconstruction is frequently performed using various equations. Some of them are listed below:

1) Perspective camera projection:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{matrix} R \\ 0^T \end{matrix}\right. \qquad (3)$$

where LHS: the pixel coordinates of the 2D image observation,
RHS 3rd matrix: are the 3D coordinates of the point in world coordinates,
, : are the focal lengths,
, : being the principal point coordinates,
R: being the rotation matrix, and
t: being the translation vector.

2) Epipolar constraint:

$$Fx = 0 \qquad (4)$$

where x and x' are respective observations of 2D image from two different views, and F is the fundamental matrix which describes the relationship between the two views.

One of SFM's key benefits is its ability to create extremely detailed 3D models using just a camera or combination of cameras. Furthermore, as long as they can be seen in the input photos, SFM is capable of handling arbitrary object shapes and surface textures. SFM is widely applicable in a variety of disciplines, such as computer vision (CV), robotics, and remote sensing. It has applications in the mapping of expansive outdoor spaces for autonomous vehicles or drones as well as the 3D modeling of buildings and monuments and the reconstruction of natural scenes for virtual reality or gaming.

E. MVS

A technique called Multi-View Stereo (MVS) allows you to piece together the 3D geometry of objects or scenes using a variety of 2D images that were shot from different perspectives. The method involves finding the 3D coordinates of points on an object's surface by comparing the relationships between points in various photos.

The method calculates the camera poses, or the position as well as orientation of the camera with respect to the object, for each input image before performing MVS. Techniques like structure from motion or calibration targets can be used to accomplish this. The method involves finding the 3D coordinates of points on an object's surface by comparing the relationships between points in various photos.

Because it makes use of every image that is available, MVS differs from other methods like SfM and SfM-based stereo in that it estimates the depth of every point in the scene. As a result, it generates 3D reconstructions that are more precise and detailed. As long as they are discernible in the input photos, MVS can also handle intricate item shapes and surface textures. Most Multi-View Stereo (MVS) techniques include minimizing an energy function made up of a regularization term and a data term. The regularization term encourages smoothness in the estimated depth map, while the data term evaluates similarity between the observed pictures and the reconstructed depth map. The energy function for MVS can be expressed in the following general form:

$$E(D) = E_{data(D)} + lambda * \quad (5)$$

where D is the depth map, lambda is a regularization parameter, and and are the data term and regularization term, respectively.

The data term is typically defined as:

$$E_{data(D)} = sum_i(sum_x(sum_y(w_{i(x,y)} * abs(I_{i(x,y)})))) \quad (6)$$

where i indexes over the input images, (x, y) are pixel coordinates, = i-th image's intensity at (x,y), is the intensity value of the warped i-th image at (x, y) using the estimated depth map D, and is a weight function that can be used to assign different weights to different images based on their quality or other criteria.

Typically, the regularization term has the following definition:

$$E_{reg(D)} = sum_x(sum_y(|grad_x(D(x,y))| + |gr \quad (7)$$

where grad x and grad y are denoted as the gradients in the x and y directions, respectively, and p controls how smooth the predicted depth map is.

Utilizing optimization methods such as Graph Cut, Belief Propagation, or Variational Techniques, the energy function is reduced to produce the final depth map.

IV. RESULT AND DISCUSSION

For MVS, the authors in [8] utilized a dataset of 61 images of a statue from different angles. The images were fed as input to the algorithm and as an output it reconstructed a 3D model.

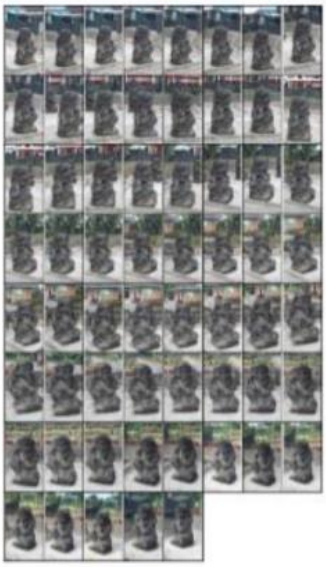

Figure 2.Image dataset.

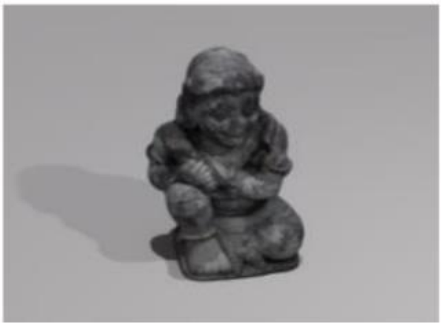

Figure 3.3D Model reconstruction.

DOI: https://doi.org/10.17762/ijritcc.v11i6s.6942
***Article Received**: 26 March 2023 **Revised**: 18 May 2023 **Accepted**: 28 May 2023*

When SFM is considered, One view must be regarded as camera 1 and the other as camera 2 for the straightforward situation of structure from two fixed cameras or one moving camera. Camera 1 is positioned at the origin, and the algorithm in this case assumes that its optical axis is parallel to the z-axis.

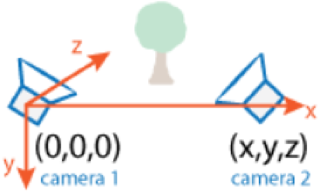

Figure 4. Position of cameras shown on axis

In [9], following process was followed to successfully implement the algorithm of SFM:

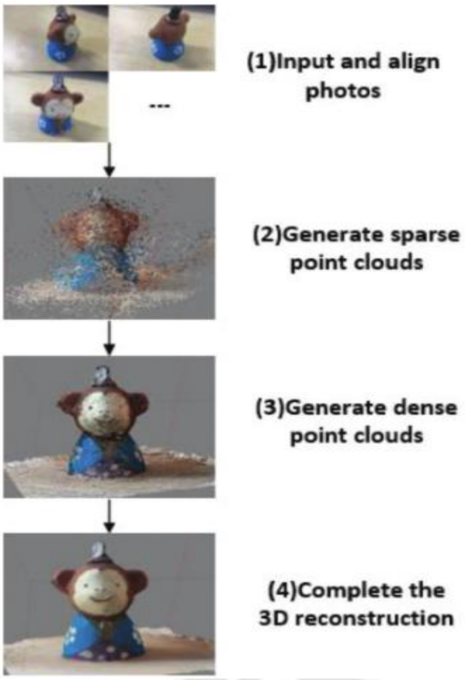
(1)Input and align photos
---
(2)Generate sparse point clouds
(3)Generate dense point clouds
(4)Complete the 3D reconstruction

Figure 5.Steps for generating SFM output.

Moving on to SFS, this approach is comparatively primitive. The SFS model presupposes that the surface normal, albedo, and light source direction are functions of the measured picture intensity. With the help of illumination and shading data that are already known, SFS attempts to estimate the unknown surface normal. The complexity of SFS makes it a difficult problem to solve, and a number of variables can affect how accurately the surface normal is computed. The output of this algorithm is demonstrated below.

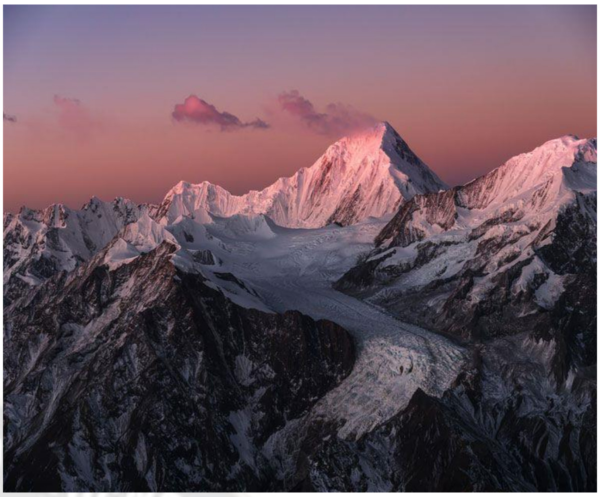

Figure 6. (Input) Image given.

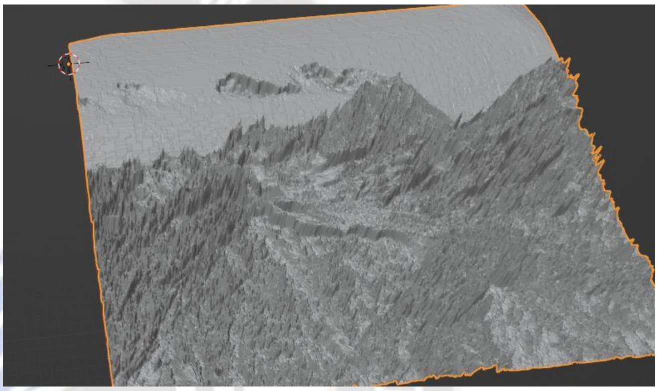

Figure 7. (Output) Image constructed.

As it can be observed from the above images, the 3D model reconstruction from SFS algorithm is not very accurate. When compared to PIFu, the Structured from Motion (SfM) and Shape from Shading (SfS) techniques fall short in terms of producing precise 3D models. SfM makes it challenging to obtain dense 3D models since it takes a lot of photos and a lot of overlap between them. On the other hand, SfS has a limited ability to recreate surfaces with fine features or texture and is highly dependent on lighting conditions. Contrarily, PIFu makes use of a neural network to infer 3D shapes from 2D photographs, making it more reliable and accurate at creating complex 3D models even from a single image.

The last one is PIFu and PIFu hd. PIFu has a number of benefits over other 2D to 3D reconstruction algorithms, including:

- Increased precision: PIFu utilizes a deep neural network to learn the object’s geometry and estimate the surface normal at each pixel, leading to more precise and complete 3D reconstructions.
- Performance in real-time: PIFu can produce 3D models in real-time, making it appropriate for uses like virtual try-on and augmented reality.

- PIFu is more generalizable than other algorithms since it can create 3D models of things with variable shapes and textures, as opposed to other methods that depend on precise assumptions about the geometry of the object. PIFu is simple to use and takes little input from the user, making it usable by non-experts.

In general, PIFu significantly outperforms current 2D-to-3D conversion algorithms by means of accuracy, real-time performance, and generalizability.

As presented in [1], following are the results obtained from implementing PIFu:

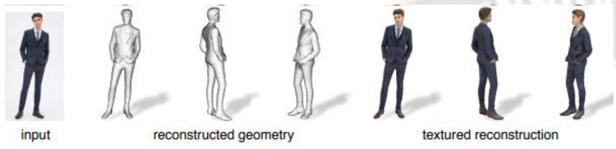

Figure 8. Input and output of PIFu

A better version of PIFu, called PIFu HD, was created with high-resolution image inputs in mind. Their network designs are the primary distinction between PIFu and PIFu HD. With additional convolutional layers and a better quality feature map than PIFu, PIFu HD employs a deeper network. Furthermore, PIFu HD employs a multi-scale strategy to more effectively gather features at various scales.

PIFu HD, as opposed to PIFu, generates 3D reconstructions of higher quality and with more minute features. Due to its more complex network architecture, PIFu HD is slower than PIFu and needs more processing resources. The exact application and the available computational resources will determine which of PIFu and PIFu HD to use.

The above discussed algorithms are only a few of many algorithms that can be utilized to implement conversion of 2-D image to 3-D model. We reviewed around 40 different papers on this topic and summarized their approach and findings in table 1. Table 1 contains 5 main columns. First is the year of publication of paper, the method proposed or used, dataset used by them to implement their methodology or algorithm, the performance metrics used by the authors of the papers to analyze their results and measure their success and the last is values of those performance metrics. Table 1 gives concise information about all the papers reviewed and helps understand the current scene or trend in the industry.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6s
DOI: https://doi.org/10.17762/ijritcc.v11i6s.6942
***Article Received**: 26 March 2023 **Revised**: 18 May 2023 **Accepted**: 28 May 2023*

TABLE 1.
Observations on Results, methods, and Datasets.

P. ID	Year	Method Used	Dataset Used	Performance Matrices	Value
[1]	2019	PIFu	Buff	Chamfer	1.14
[2]	2019	Pix2Vox-A	ShapeNet,Pix3D	IoU	0.288
[3]	2020	ML-PIFu (alternate) PIFuHD ML-PIFu (end-to-end)	BUFF CAPE RenderPeople	Chamfer	1.1.73 2.3.237 3.1.525
[4]	2021	Generalized Binary Search Network (GBiNet)	DTU	Accuracy	0.327
[5]	2022	Shape Encoder	Market-HQ	MaskIoU (%)	81.1
[6]	2022	ONet network architecture model(proposed method)	3DPeople	IoU Chamfer	1.0.610 2.0.100
[6]	2022	3D-C2FT	ShapeNet	F-score	0.464
[7]	2022	FvOR-Pose	ShapeNet	Pixel Error	18.0/5.0
[8]	2022	View-dependent Depth Sampling	ShapeNet-S, ProSketch and AmateurSketch	Chamfer Distance	9.515 3.868 9.657
[9]	2022	SimpleRecon	ScanNetv2	F-score	0.671
[10]	2022	SSP3D	ShapeNet Pix3D	mean IoU	61.64 35.39
[11]	2022	PIFuHD PIFu	NeRF	CD	0.099† 0.048
[12]	2022	Multi-View Guided MVS	BlendedMVG DTU	Accuracy	0.339 mm 0.325
[13]	2022	Non-metric Multi - Dimensional scaling (NMDS)	CelebA CASIA3D-	Average MSE	0.14 ± 0.037 1.50 0.18
[14]	2022	EG3D-RenderDiffusion	ShapeNet Clevr	PSNR	25.4 39.8
[15]	2022	Bootstrapped Radiance Field Inversion	SRN Cars,SRN Chairs ,CARLA	Pose estimation accuracy	10.84◦ 7.29◦ 1.08◦
[16]	2022	DP-NeRF	synthetic and real scene dataset	PSNR SSIM LPIPS	23.67 0.7299 0.1082
[17]	2022	BAD-NeRF	Deblur-NeRF Synthetic and Real datasets	PSNR SSIM LPIPS	29.3 0.87 0.11
[18]	2022	Fast-SNARF	DFaust	Accuracy	81.2%
[19]	2022	PatchMatch-Stereo-types	Self Created dataset	SLAM densification	1.119s, 2.+0
[20]	2022	NeuralLift-360	NeRF	CLIP Distance	0.4498
[21]	2022	NOPE-SAC	Matterport3D,ScanNet	Translation mean	1.0.94, 2.0.65

***Article Received**: 26 March 2023 **Revised**: 18 May 2023 **Accepted**: 28 May 2023*

[22]	2022	NeRDi	DTU	PSNR	8.000
[23]	2022	SceneRF	SemanticKITTI BundleFusion	Recall	1.40.96 2.47.92
[24]	2022	SDFusion	ShapeNet BuildingNet Pix3D	fidelity (UHD) diversity (TMD) Chamfer Distance F-Score	0.0557, 0.1116 0.0885, 0.0745 3.1.852 4. 0.432
[25]	2023	Mono-STAR	FlyingThings3D Monkaa Sintel	SAD-graph	t=38
[26]	2023	MULTI-VIEW SURFACE RECONSTRUCTION,D COLOUR-GUIDED DEPTH MAP ENHANCEMENT	MobileBrick	Accuracy	85.9%
[27]	2023	Multiview Compressive Coding (MCC)	Hypersim	Accuracy	66.3%
[28]	2023	Attention Collaboration-based Regressor	InterHand2.6M, FreiHand	MPJPE	1.7.41 2.6.9
[29]	2023	Multi-Person Coarse Reconstruction	Portrait Relighting Dataset (PR-Senior, PR-Young)	PSNR	1.30.60 2.30.13
[30]	2023	Zero-1-to-3	Google Scanned Objects RTMV	PSNR	1.18.378 2.10.405
[31]	2023	NeuS+	NeRF	Chamfer-L1	22.07
[32]	2023	Zolly	PDHuman,SPEC-MT HuMMan	PA-MPJPE	1.39.4 2.65.8 3.23.0
[33]	2023	PAniC-3D	Vroid 3D dataset Vtuberdataset AnimeRecon benchmark	LPIPS CLIP PSNR	1. 18.26 2. 94.97 3. 16.96
[34]	2023	MVAS	Real life objects	1.Chamfer distance 2. F-score	1.0.307 2.0.816
[35]	2023	SVIn2	GinnieBallroom Cenote Coral Reef	Precision	1.85.6 2.91.2 3.81.8
[36]	2023	ALIKE-N	Hpatches	MMA MHA	75.23% 74.44%
[37]	2023	Progressive Volume Distillation with Active Learning	Synthetic-NeRF,LLFF TanksAndTemples	R2L+AL KiloNeRF+AL	30.35 29.21
[38]	2023	1.JaxNeRF 2.Plenoxels 3.DVGO	DTU	IMRC	18.54 16.00 18.27
[39]	2023	1.Apparent Contour Event (ACE) 2. Mask-24 3. Mask-12	MOEC-3D	Chamfer Distance	2.4267 3.2652 4.3856
[40]	2023	Simple Learned Keypoints(SiLK)	ScanNet	Accuracy	99.1%

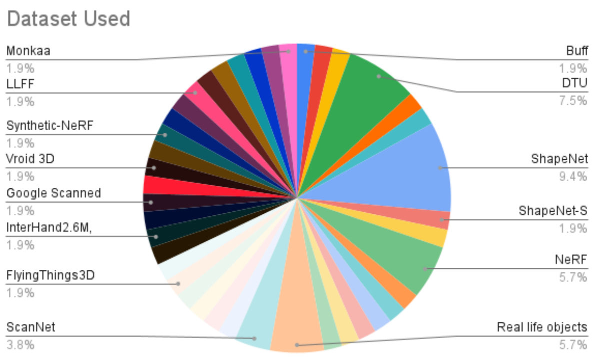

Figure 9. Datasets encountered in the survey of all papers

The pie chart in Figure 9 represents all the various datasets that have been used in the papers that have been listed in Table 1. As observed most datasets that was used is ShapeNet dataset, DTU being a close second. This could be because these are standard datasets, which are usually used for experimentation or implementation. The third most used dataset is a tie between NeRF standard dataset and real life object dataset.

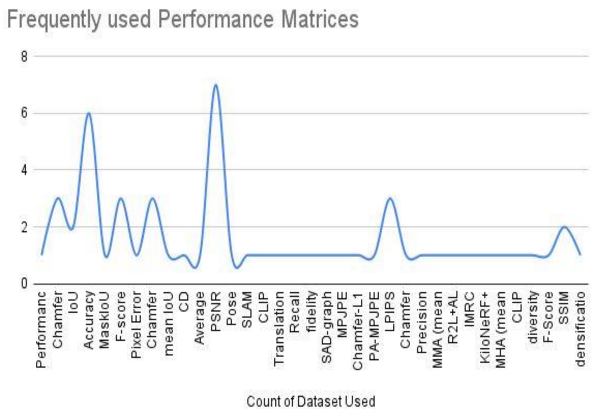

Figure 10. Performance metrics used in all papers

Table 1 also lists the performance matrices that have been used by the respective authors to compare and analyze the efficiency and overall performance of their implemented algorithm. The graph in Figure 10, shows the most used performance metrics. It is PSNR, Accuracy being a close second. The peak signal-to-noise ratio between two pictures is calculated by the PSNR block, it is expressed in decibels. The quality of the original and compressed photos is contrasted using this ratio.

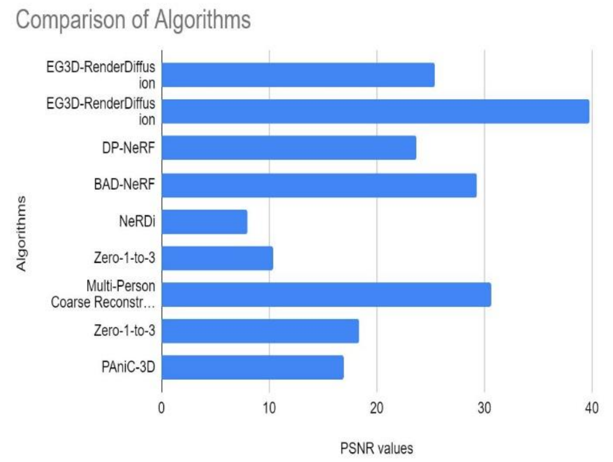

Figure 11. Comparing PSNR values all algorithms

The bar graph in Figure 11, shows comparison of PSNR values of various algorithms implemented in Table 1. It can be inferred that EG3D-RenderDiffusion is best performing. Multi-Person Coarse Reconstruction is second best.

A. Algorithm 1 Feature extraction from a image

```
function extractFeatures(image):
    keypoints = detectKeypoints(image)
    descriptors = computeDescriptors(image, keypoints)
    return keypoints, descriptors
function detectKeypoints(image):
    // Use a feature detection algorithm
    //(e.g. FAST, Harris, ORB) to detect keypoints in the image)
    detector = FeatureDetector.create()
    keypoints = detector.detect(image)
    return keypoints
function computeDescriptors(image, keypoints):
    // Use a descriptor extraction algorithm (e.g. SIFT, SURF, ORB)
    //to compute feature descriptors for each keypoint in the image
    descriptorExtractor = DescriptorExtractor.create()
    descriptors = descriptorExtractor.compute(image, keypoints)
    return descriptors
```

This technique takes an input image and outputs a list of key points that were found together with their corresponding descriptors. A feature detection approach is used by the detectKeypoints() function to locate distinctive areas or points in the image. The computeDescriptors() method uses a descriptor extraction algorithm to compute feature descriptors for each key point. In a later stage of Vuforia's feature-based image mapping technique, these descriptors are utilized to match the key points with reference images.

B. Algorithm 2 Image feature mapping

```
function matchFeatures(keypoints, descriptors,
referenceImage):
referenceKeypoints, referenceDescriptors =
getReferenceFeatures(referenceImage)
   matcher = FeatureMatcher.create()
   matches = matcher.match(descriptors,
referenceDescriptors)
filteredMatches = filterMatches(matches)
homography=computeHomography(keypoints,referenc
eKeypoints, filteredMatches)
   return homography

function getReferenceFeatures(referenceImage):
referenceKeypoints, referenceDescriptors =
extractFeatures(referenceImage)
   return referenceKeypoints, referenceDescriptors

function filterMatches(matches):
   // Apply a filtering method (e.g. ratio test,
   //geometric consistency) to remove unreliable
matches
filteredMatches = []
   for match in matches:
      if isMatchValid(match):
filteredMatches.append(match)
   return filteredMatches

function computeHomography(keypoints,
referenceKeypoints, matches):
   // Use a homography estimation algorithm (e.g.
RANSAC, LMEDS)
  //to compute the transformation matrix between the
  // keypoints and the reference keypoints
homography, _ = findHomography(keypoints,
referenceKeypoints, matches)
   return homography
```

This algorithm outputs the homography matrix, which defines the transformation between the two pictures, after receiving as inputs the keypoints and descriptors of an image and a reference image. The getReferenceFeatures() function collects the keypoints and descriptors from the reference image while the matchFeatures() function compares the keypoints and descriptors of the input image with those of the reference picture. A homography matrix is constructed using the filtered matches, the related keypoints, and the reference keypoints after the matches have been filtered to remove unreliable matches. An augmented reality experience can be made by superimposing virtual objects on the input image using the homography matrix.

C. Algorithm 3 NeRF

```
initialize the input image
initialize the neural network model
initialize the optimization algorithm
initialize the 3D volume

while true do:
   feed the input image to the neural network
   obtain the 3D volume from the neural network
output
   optimize the 3D volume using the optimization
algorithm
   if need to visualize 3D model:
      render the 3D volume using a volume rendering
library
   update the input image (if desired)
```

Using NeRF, this program converts a 2D image into a 3D volume. The neural network outputs a 3D volume using the input image as its input, represented as a grid of voxels or another kind of spatial representation. This 3D volume represents the scene shown in the image taken as input. To match the 3D volume to the input image, an optimization approach (such as gradient descent) can be used. Other restrictions on the scene's form and appearance may also be included in the optimization, along with phrases that encourage the volume to match the image taken as input.

The user can view the 3D model of the scene shown in the input image after optimization by using a volume rendering library to see the 3D volume. To create fresh or enhanced 3D models, the algorithm can be run again with an updated input image.

D. Algorithm 4 PiFu HD

```
initialize the input image
initialize the neural network model
initialize the optimization algorithm
initialize the 3D mesh

while true do:
   feed the input image to the neural network
   obtain the initial 3D mesh from the neural network
output
   optimize the 3D mesh using the optimization
algorithm
   refine the 3D mesh using Laplacian smoothing and
other techniques
```

```
if need to visualize 3D model:
    render the 3D mesh using a 3D rendering library

update the input image (if desired)
```

Using PiFu HD, this program converts a 2D image into a 3D model. The neural network generates a first 3D mesh as its output using the input image as its input. To fit the 3D surface to the input image, this mesh is then optimized using an optimization process (such gradient descent). The optimization can incorporate conditions that promote 3D surface smoothness, conformance to the input image, and other restrictions. Following optimization, the 3D mesh is enhanced using methods such Laplacian smoothing to raise its quality. By visualizing the 3D mesh using a 3D rendering library, the user can visualize the object’s 3D model depicted in the image taken as input. To create fresh or enhanced 3D models, the algorithm can be run again with an updated input image.

V. CONCLUSION

This paper's main goal was to provide a succinct overview and assessment of the current technologies and approaches used in the field of converting 2-D images to 3-D models. In the paper, pertinent methods are covered, including SFS, SFM, MVS, PIFu, and PIFu HD. It also makes an effort to thoroughly explain these algorithms. It goes on to explore some of the techniques' drawbacks and comes to the conclusion that PIFu HD is the most recent and effective method for carrying out the conversion of a 2-D image into a 3-D model.

REFERENCES

[1] Natsume, R., Morishima, S., Kanazawa, A., and Li, H. Saito et al., 2019. Pifu: Implicit pixel-aligned function for high-resolution human digitisation in clothing. (Pp. 2304–2314) in Proceedings of the IEEE/CVF International Conference on Computer Vision.

[2] Chan, E.R. Lin, C.Z Chan, M.A. Nagano, K. Pan, B. De Mello, S. Gallo, L.J. Guibas, J. Tremblay, S. Khamis, and T. Karras, 2022. 3D generative adversarial networks that are effective and mindful of geometry. pp. 16123–16133 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[3] Saito, S., Simon, T., Saragih, and Joo. Pifuhd: Multi-level implicit function with pixel alignment for high-resolution 3D human digitisation. On pages 84–93 of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[4] A. Kanchan and T. Mathur, 2017. Recent developments in 2D to 3D picture conversion: a quick look at the algorithm. pp. 3480–3484 in International Research Journal of Engineering and Technology, 4(4).

[5] S. McCann, 2015. 3D reconstruction using several pictures. Geometric and Computational Vision Lab at Stanford.

[6] 2014; Victoria M. Baretto International Journal of Engineering Research and Technology (IJERT) NCRTS - 2014 (Volume 2 - Issue 13), Automatic Learning based 2D-to-3D Image Conversion.

[7] Wei, X., Zhang, Y., Li, Z., Fu, Y., and Xue, X. Structure from motion using deep bundle adjustment, or deepsfm. 16th European Conference on Computer Vision, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 (pp. 230–247). International Springer Publishing.

[8] Moch, Kholil & Ismanto, I & Fu’ad, M. (2021). “3D reconstruction using Structure From Motion (SFM) algorithm and Multi View Stereo (MVS) based on computer vision.” ,presented in IOP Conference Series: Materials Science and Engineering. 1073. 012066. 10.1088/1757-899X/1073/1/012066.

[9] Li and Jiwu, Chenyang and Yi, Shijie, and Li, Zixin, et al. (2019).Modeling Virtual Reality Environment Based on SFM Method . 6.179.10.2991/jrnal.k.191202.007 in Journal of Robotics, Networking, and Artificial Life.

[10] Multi-view stereo in the "Deep Learning Era: A Comprehensive Review," Displays, Volume 70,2021, 102102, ISSN 0141-9382, by Xiang Wang, Chen Wang, Bing Liu, Xiaoqing Zhou, Liang Zhang, Jin Zheng, and Xiao Bai,
https://doi.org/10.1016/j.displa.2021.102102.

[11] P. Gleize, W. Wang, and M. Feiszli, "SiLK--Simple Learned Keypoints" Print accessed at arXiv:2304.06194, 2023.

[12] Wang, Z., K. Chaney, and K. Daniilidis, November 2022. "EvAC3D: From Event-Based Apparent Contours to 3D Models via Continuous Visual Hulls" 17th European Conference on Computer Vision, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII (pp. 284–299).

[13] Cham.Fang, Q., Song, Y., Li, K., Shen, L., Wu, H., Xiong, G. and Bo, L., 2023. Evaluate Geometry of Radiance Field with Low-frequency Color Prior. arXiv preprint arXiv:2304.04351.

[14] S. Fang, Y. Wang, Y. Yang, X. Wu, H. Wang, W. Ding, and S. Zhou, 2023. "PVD-AL: Progressive Volume Distillation with Active Learning for Efficient Conversion Between Different NeRF Architectures". Print accessed at arXiv:2304.04012.

[15] [14] Zhao, X., Wu, X., Chen, W., Chen, PC, Xu, Q, and Li, 2023. The phrase "ALIKED: A Lighter Keypoint and Descriptor Extraction Network via Deformable Transformation" Preprint available at- arXiv:2304.03608.

[16] [15] Cao, X. Santo, F. Okura, and Y. Matsushita, 2023. The phrase "Multi-View Azimuth Stereo via Tangent Space Consistency" Preprint available at- arXiv:2303.16447.

[17] Wang, W., Joshi, B., Burgdorfer, N., Batsos, K., Li, A.Q., Mordohai, P., and Rekleitis, 2023. The phrase "Real-Time Dense 3D Mapping of Underwater Environments" preprint available at- arXiv:2304.02704.

[18] An, S., chen, S., zhang, K., Shi, Y., Zhu, Y., Wang, H., Song, G., Kristjansson, J., Yang, X. and Zwicker, M., 2023. “PAniC-3D: Stylized Single-view 3D Reconstruction from Portraits of Anime Characters”. arXiv preprint available at- arXiv:2303.14587.

[19] Sun, Q., wang, W., Ge, Y., Mei, H., Cai, Z., et, Al. 2023. “Zolly: Zoom Focal Length Correctly for Perspective-Distorted Human Mesh Reconstruction”. arXiv preprint at- arXiv:2303.13796.

[20] tong, J., muthu, S., et. Al. (2023). “Seeing Through the Glass: Neural 3D Reconstruction of Object Inside a Transparent Container”. arXiv preprint at- arXiv:2303.13805.

[21] Wu, R., Liu, R., Tokmakov, Van Hoorick, B.,P., Zakharov, S. and Vondrick, C., (2023).” Zero-1-to-3: Zero-shot One Image to 3D Object”. arXiv preprint available at- arXiv:2303.11328.

[22] zhang, J., Xu, B., He, Y., Qian, C. and Lin, K.Y., (2023). “Deformable Model Driven Neural Rendering for High-fidelity 3D Reconstruction of Human Heads Under Low-View Settings”. arXiv preprint, can be accessed at-arXiv:2303.13855.

[23] Yu, Z., Huang, Fang, Breckon, et. Al. (2023). “ACR: Attention Collaboration-based Regressor for Arbitrary Two-Hand Reconstruction”. arXiv preprint is available at - arXiv:2303.05938.

[24] Johnson, J., Wu, Feichtenhofer, C. Malik, J., et, Al. (2023). “Multiview Compressive Coding for 3D Reconstruction”. arXiv preprint paper available here - arXiv:2301.08247.

[25] castle, R., Li, bian, J.W., torr, P.H. and Prisacariu, V.A., (2023). “MobileBrick: Building LEGO for 3D Reconstruction on Mobile Devices”. arXiv preprint available here-arXiv:2303.01932.

[26] Boularias, A., Ramesh, Gan, Y. D.M., Geng, S., and chang, H. “Mono-STAR: Mono-camera Scene-level Tracking and Reconstruction”.(2023)arXiv access preprint here - arXiv:2301.13244.

[27] cheng, Y.C., tulyakov, S., Lee, H.Y., Schwing, A. and gui, L. “SDFusion: Multimodal 3D Shape Completion, Reconstruction, and Generation”.(2022) arXiv preprint available at-arXiv:2212.04493.

[28] Guibas, L., jiang, C., Qi, C.R., yan, X., zhou, Y, Deng, C. and Anguelov, D. “NeRDi: Single-View NeRF Synthesis with Language-Guided Diffusion as General Image Priors”. (2022). arXiv preprint available at- arXiv:2212.03267.

[29] de Charette, R., and Cao, A.Q. “SceneRF: Self-Supervised Monocular 3D Scene Reconstruction with Radiance Fields”. (2022). arXiv preprint can be accessed via- arXiv:2212.02501.

[30] Tan, Wu, T. xue, N., and xia, G.S. “NOPE-SAC: Neural One-Plane RANSAC for Sparse-View Planar 3D Reconstruction”.(2022). arXiv preprint available here-arXiv:2211.16799.

[31] Jiang, Y., Xu, D., et. Al. “NeuralLift-360: Lifting An In-the-wild 2D Photo to A 3D Object with 360° Views”.(2022). arXiv e-prints, available here- pp.arXiv-2211.

[32] surmann, H., et. Al.. PatchMatch-Stereo-Panorama, a fast dense reconstruction from 360 {\deg} video images”.(2022) arXiv preprint available at - arXiv:2211.16266.

[33] X. Chen, T. Jiang, J. Song, M. Rietmann, A. Geiger, M.J. Black, and O. Hilliges, 2022. "Fast-SNARF: A Neural Field Deformer for Articulated Neural Fields". arXiv:2211.15601 is an arXiv preprint.

[34] P. Wang, L. Zhao, R. Ma, and P. Liu. "BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields",(2022). arXiv preprint: 2211.12853.

[35] Lee, D., et. Al. (2022). “Deblurred Neural Radiance Field with Physical Scene Priors”. arXiv preprint can be accessed at-arXiv:2211.12046.

[36] Anciukevičius, T., et. Al. (2022). “RenderDiffusion: Image Diffusion for 3D Reconstruction, Inpainting and Generation”. arXiv preprint can be accessed at- arXiv:2211.09869.

[37] Tombari, F., pavllo, D., rakotosaona, M.J. and tan, D.J., (2022). “Shape, Pose, and Appearance from a Single Image via Bootstrapped Radiance Field Inversion”. arXiv preprint available here- arXiv:2211.11674.

[38] Azimifar, Z., along with Kamyab, S., (2022). “Deep-MDS Framework for Recovering the 3D Shape of 2D Landmarks from a Single Image”. arXiv preprint available here-arXiv:2210.15200.

[39] Poggi, M., et. Al. (2022), October. “Multi-View Guided Multi-View Stereo”. Published in - IEEE/RSJ International Conference on Intelligent Robots and Systems .IEEE.

[40] Koulieris, G.A., chang, Z., and shum, H.P., (November 2022). “3D Reconstruction of Sculptures from Single Images via Unsupervised Domain Adaptation on Implicit Models”. Publiahed in Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (pp. 1-10).

[41] Jiang, Y.G.,., li, H., wu, Z. and xing, Z . “Semi-supervised Single-View 3D Reconstruction via Prototype Shape Priors”. Published in Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, (October 23–27, 2022), Proceedings, Part I (pp. 535-551).

[42] Gibson, J., Sayed, M., et.Al. “SimpleRecon: 3D reconstruction without 3D convolutions”. Published In Computer Vision–ECCV 2022: European Conference(17th edition), Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII (pp. 1-19)

[43] Sun, X., Zhou, S., and Zhang, S., Xie, H., Yao, H., Sun, X., and Zhou, S.(2019). "Pix2vox: Context-aware 3d reconstruction from single and multi-view images" (Pp. 2690–2698) in Proceedings of the IEEE/CVF International Conference on Computer Vision.

[44] Gao, C., Yu, Q., Sheng, L., Song, Y.Z., and Xu, D. "SketchSampler: Sketch-Based 3D Reconstruction via View-Dependent Depth Sampling". European Conference on Computer Vision(17th edition), Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part I (pp. 464–479). Springer Nature Switzerland, Cham.

[45] Ayaz , M. ., & Ammad Uddin , M. . (2023). Performance Analysis of Underwater Wireless Optical Communication with Varying Salinity: Experimental Study. International Journal of Intelligent Systems and Applications in Engineering, 11(1), 18–24. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2439

[46] "Generalised binary search network for highly-efficient multi-view stereo" by Mi, Z., Di, C., and Xu, D., 2022. (Pages 12991–13000) in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[47] Yang, Z., et. Al. (2022). “HM3D-ABO: A Photo-realistic Dataset for Object-centric Multi-view 3D Reconstruction”. arXiv preprint available at -arXiv:2206.12356.

[48] Pumarola, A., ugrinovic, N., et. Al. “Single-view 3D Body and Cloth Reconstruction under Complex Poses”. (2022) arXiv preprint can be accessed here- arXiv:2205.04087.

[49] Sigmund, D. tiong, L.C.O., and teoh, A.B.J. “3D-C2FT: Coarse-to-fine Transformer for Multi-view 3D Reconstruction”.

(2022)Printed In Proceedings of the Asian Conference on Computer Vision (pp. 1438-1454).

[50] Yang, Y, Z. Zheng, J. Zhu, Ji, W., and chua, T.S."3D Magic Mirror: Clothing Reconstruction from a Single Image via a Causal Perspective".(2022). arXiv preprint can be accessed via-arXiv:2204.13096.

[51] Gui Shengxi et. Al. "Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto", (2021)Proceedings of "ISPRS Journal of Photogrammetry and Remote Sensing", Volume 181, Pages 1-19, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2021.08.025.

[52] Upadhye Gopal & Kulkarni U.V. & Mane Deepak, "Improved Model Configuration Strategies for Kannada Handwritten Numeral Recognition" Image Analysis & Stereology. Vol.40, pp.181-191. 10.5566/ias.2586