ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230518000646
Original scientific paper

Fuzzy-Logic based Effective Contour Representation of Occluded Objects

H. SHALMA*, P. SELVARAJ

Abstract: We present a fuzzy-based network for the sharpening of object contour even in the presence of occlusion. The contour representation of objects can be effectively handled by the structure tensor method. This work proposes an occlusion detection and filling strategy using the square patch selection method. Based on the interpolation method, the fuzzy-assisted square patch selection can be used to fill the occluded pixels. Due to the occluded pixels, the depth map may have anomalies in the low-texture and high-exposure areas. Before converting a depth map to a point cloud, it is essential to filter out the outliers in the depth map to obtain a more accurate point cloud. To improve the precision of the depth map, improved occlusion detection and management procedures is required.The occlusion regions may be confirmed through belief propagation, which may produce noisy results in occluded regions, sharp objects, and object boundaries. We strived to build a model that differentiates the occluded pixels from others by exploiting sharp boundary transitions. We have used a stereo geometry structure to develop the required deep neural models to handle occlusion. We built the model by creating layers for every pipeline component and made it to learn the contour representation model using an adaptive fuzzy-based approach. In existing approaches, the bias must be properly predicted with the Gaussian distribution. The proposed model eradicated the pixel bleeding effect by exploiting the bimodal distribution with Gaussian and SMD (Stereo Mixture Density) functions and by finding smoothening bias.The suitable depth values were assigned to the occluded regions obtained. The experimental results demonstrated that the proposed approach generates more stable depth maps with fewer constraints than the existing methods. The experimental results were compared with the standard SMD-Net and other state-of-the-art models.

Keywords: contour representation; deep learning; fuzzy decision; occlusion filling; sharp boundary

1 INTRODUCTION

In computer vision, estimating a depth for 2D to 3D image conversion is challenging. Identifying sharp boundaries in a cluttered/non-cluttered environment is also difficult. Despite stunning results, predicting correct depth bounds and generating super-resolution outputs with limited memory and processing time remains challenging. Neural Networks recreate object boundaries with "bleeding/distractive" pixels in point clouds. These artifacts might severely hinder 3D reconstruction and object detection. The pixel variations among the point cloud and the ground truth values would result in depth discontinuities. The amount of depth discontinuity could impact the level of sparsity. Such sparse alignment of depth information would result in poor object elevation [1]. The depth maps must be computed in such a way as to avoid creating any pixel bleeding across the object's boundaries for dense object reconstruction [2]. Conventional methods for object boundary prediction do not rely on patch centers. Adding up-sampling layers to boost output resolution requires a lot of memory and computing [3]. The Boundary bleeding artifacts may occur because of the disparities in in-depth map [4] prediction. An edge-preserving usually applies the grab-cut method, canny operators, and structure tensors are used for effective object border discrimination. Many researchers have applied fuzzy logic in edge detection, image segmentation, and object recolonization. Depth discontinuities can be predicted through a fuzzy-based structure tensor method. We include adaptive fuzzy-based filters to handle such cases for better edge preservation in object boundary structures.The bimodal distribution can correct the bleeding effect in 3D contour reconstruction for the generated disparity maps. Another distorting effect, anisotropic diffusion, appears in crossing/overlapping edges and can be corrected using a guided filter. The heuristics are usually preferred to correct the edges with less filling and computational time. We considered only the occluded patches for the filling process. In this work, we inscribe two issues. Our contribution aims to learn a precise representation at object borders even when occlusion is present. The adaptive square patch selection method was proposed to combat the pixel bleeding artifacts arising during the conversion while converting the depth map to the point cloud format. An algorithm was proposed and guided filter used to identify the valid disparity values and fill the disparities through the neural network. The rest of the paper is organized as follows: Section 2 describes existing work about object contour representation; section 3 addresses the proposed methodology of adaptive fuzzy-based guided filter for effective object reconstruction; section 4 addresses a neuro based guided filtering process for object contour representation; section 5 describes the Experiments and results, and section 6 concludes overall work.

2 RELATED WORKS

The disparities of the stereo images R. A. Hamzah and H. Ibrahim [5] described the edge detection operations using Gaussian (DoG) differentiation. The decomposition of layers in pyramid levels requires a multi-scale edge fusion technique that aids in fusing the different-scale maps in each phase. Medeiros, Marcos D, et al. [6] described the role of fuzzy decision and its role in extracting local features from pair of images using correspondence points. The trading between the error and time has been solved using a fuzzy approach by exploiting the various resolution levels. In [7], a neural net-based approach DispNetC was proposed by exploiting stereo cost volume and Depth Image-based Rendering (DIBR)[8]. The hole-filling strategy with the Disparity approach fills small or medium-sized holes based on categorization while maintaining high image quality. If both of a hole's bordering pixels are foreground pixels, then the pixel mismatches might create the overall visual quality of the patched holes painted over [9]. Zbontar et al. [10] proposed an approach to estimate the discrepancies with a Siamese network to determine potential partners' distance from picture squares. To quantify the gap, the authors then aggregate costs across multiple units of analysis [11] and SGM (Semi-Global Matching) [12]. This method does not train a convolutional network for the disparity estimation task from start to finish, which affects both computing efficiency and accuracy. Liu, Wei Chen, et al. [13] described how the Semi global Weighted Least Square method (SG-WLS) improves space and time complexity. They used 1D filters with the two-dimensional neighborhood information to smooth the input image and perform better. However, the neighborhood information cannot be applied directly to input sparse data. Hence they have done a guided sparse interpolation with guided depth upsampling. Zhang, Zhendong Jung, and Cheolkon [14] demonstrated an effective smoothening that preserves objects' boundaries and structure, even in the form of isotropic textures. A normalizing factor in the Gaussian kernel leads to finding the boundaries with weak gradients by finding the isotropy effect in texture regions. Hence the various approaches proposed to handle the occluded regions were discussed. The existing filtering techniques need improvement in terms of processing cost and accuracy. A novel approach with reduced processing time and improved accuracy in contour representation is highly demanded.

3 FUZZY-ASSISTED FILTERING STRATEGY

The proposed approach reduces engineering design complexity with fuzzy decisions through an adaptive-fuzzy system in selecting patches. We aim to build a 3D model with sharp boundary transitions, even in occlusion, in 3D representation. We built our model by creating layers for each stereo pipeline component. We made the system to learn the whole design with our adaptive fuzzy system and structure tensor for effective contour representation operations. The volumetric methods average the concept of semi-global block matching in occlusion handling with the extraction of feature maps at different scales by Spatial Pyramid Pooling (SPP) defined in PSMNet.

3.1 Deep Adaptive Fuzzy-Based Guided Filter Network

Our model utilizes several 2-dimensional convolutional processes to learn a deep representation. We used a convolutional filter of size 5 × 5 with a stride of two to subsample the input and lower the processing burden. Then eight residual blocks containing two series-connected 3 × 3 convolutional filters were created. The siamese architecture contains two stages of finding the similarity between the images. The input left and right images are processed to gain the image features with its similarity measure in the disparity range. Specifically, the stereo images are sent to 3 convolution layers with stride 2 for downsampling 6 resnet blocks to learn the deep features. Then, left and right image features are supplied as input to built cost volume. The cost volume is then passed to 3D CNN block, each one containing the convolution layer, batch normalization, and leaky relu activation function. Each block pertains information on disparity spatial relationship to derive the coarse disparity maps.

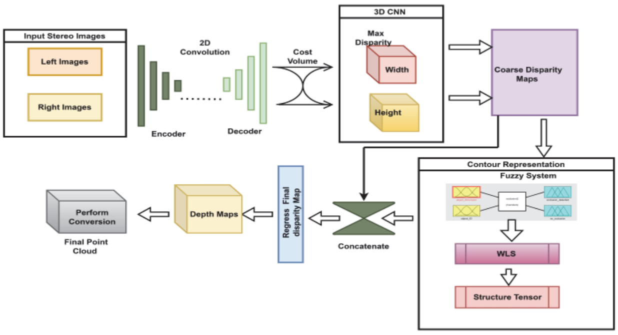

Figure 1 Architecture diagram for adaptive fuzzy-based guided filter for object reconstruction

The extra dimension in 3-D convolutions paved the way to increase the computational time required for optimization training and inference. To circumvent such issues, the deep encoder-decoder model predicts dense output employing up-sample feature maps of height, width, maxdisparity +1 by interpolation. The output of 3D CNN is a coarse disparity map fed as an input to an adaptive fuzzy system and structure tensor-based method as a function of gradient measure in contour representation. Then, an analogous refinement module to get a finer-grained estimate of the discrepancy, identifying only a residual to add or subtract from the rough prediction. The regressed disparity maps were converted to depth maps through a function of inverse proportionality to Disparity. The mixture density function of stereo with Gaussian operator is utilized for estimating the pixel bleeding effect raised in conversion from depth map to pointcloud is being corrected and produced.

3.2 Inconsistent Pixel Disparity Detection

The disparity map from the deep neural network has some depth discontinuity pixels due to the presence of outliers, and these outliers are a result of pixel mismatch or occlusion. The pixel mismatch was evaluated using cost metric, and the novel approach to tackle the occlusion problem. The original Sum of Absolute Difference (SAD) formulation used a patch of square fixed window (as correlation window) surrounding the pixels of the target image based on interest. The correlation of windows one among that moved in all places d in the second image; the lowest correlation value determined the pixel's disparity value. Each window size calculated disparity maps based on one's left image's corresponding to the right and another vice-versa. The inconsistent disparities caused by occlusions present in the scene were identified.

3.3 Adaptive Fuzzy Approach in Patch Selection

The disparity map was given as input in fuzzy set to identify occluded regions based on the categories as highly occluded, partially occluded, and no occluded. This adaptive fuzzy used a square patch to convolute the input disparity map and the Weighted Least Square (WLS) Filter. The filtering process has resulted in a smoothing effect in the inconsistent regions. The parameters involved in the filter are sigma and lambda was set to 1.5 and 8000 for the maximum disparity range of 128 in varying window sizes from 11, 13, and 31. The created disparity WLS filter is the method that utilizes all the parameters specified to obtain the smoothened output for effective contour representation. Softmax logistic regression for CNN model is used to normalize the network output described in Eq. (1) where P is the input left and right image for *i* and *j* as size.

$$\sigma(P)_i = \frac{e^{P_i}}{\sum_{j=1}^{d_{\max}} e^{P_i}} \tag{1}$$

3.3.1 Fuzzy Rule Set Algorithm for Inter-Object Occlusion

Step 1: Input RGB image converted to Grayscale. The Lightness method has reduced the contrast by taking an average of prominent values. $\frac{\{\max(R, G, B) + \min(R, G, B)\}}{2}$ Luminosity forms the weighted average $(R + G + B)/3$, for the human perception in observing events as luminosity equals $0.21R + 0.71G + 0.07B$.

Step 2: Binarize the input images with the Specified threshold denoted as $\underline{T}1$ and $\underline{T}2$.

Step 3: Generate the neighbor matrix based on the count of non-zero neighbors of each pixel for Fig. 1 and Fig. 2.

Step 4: Filling the occluded pixels is done by choosing the pixels from neighbor states with a connected neighbor technique.

Step 5: Pixel occlusion problem and divergence is solved by taking neighbors of pixel *P* in either four or eight, marking the initial pixel as *Q* alive, and the remaining un-occluded pixels are kept stable and filled using a neighborhood matrix with median value as *M*.

Step 6: If the count value of neighbor's pixel P ≤ current, and its initial state is alive, then find a pixel *Q* among four horizontal and vertical neighbors of *P*, having several neighbors equal to 1, aids missing on some edge pixels.

Step 7: Finally, the image is well-identified, even with thin edges.

The following table (Tab. 1) describes the symbols used in the proposed approach.

Table 1 List of Symbols

Symbols	Description
p	Pixel
d, d_{op_1}, (op_1)	Disparity, Disparity of occluded points
op_1, oq_1	Occluded point op and oq
w_{op_1}, w (p, d)	Occluded point weight, Filter Weights
$Cost_{grad}(p, d)$	Cost volume of gradient measure
$E(U)$	WLS Optimization result

3.4 Guided Filter for Image Smoothening

Many plane fitting algorithms, such as the RANSAC of pixels, calculate the plane coefficient. The position of the camera angle shows the parts of the object occluded in the stereo vision camera, the occluded point op_1 has the neighbors oq_1. The disparities of oq_1 is $d_{op_1}(oq_1)$ and $d'_{op_1}(op_1)$ is the Disparity of op_1 estimated from the interpolation. Based on the color similarity in the Weighted Least Square (WLS), consider that points on the foreground surface have different color intensities compared with the occlusion background.The intensity of color difference shows the inter-object occlusion.

$$\Delta = w_{opl}\left(d'_{op_1}(op_1) - d_{op_1}(oq_1)\right) \tag{2}$$

where in the Eq.(2) w_{opl} represent the weight of the occluded point. To get a refined output representation in point cloud format with sharp boundary transitions and differentiation even in inter-object occlusion, the window size chosen for the filtering process is 13 × 13. This window was made to move across the images to scan for the depth discontinuities through the fuzzy set edge detection process. The overlapping edges raised due to some interference issue also handled effectively. The edges refer to the gradient measure of pixel, and disparity.

$$cos\,t^n(p, d) = \sum_q w(p, d)\,cos\,t(p, d) \tag{3}$$

In Eq. (3) filter weights $w(p, d)$ depending on the color guidance image reduced the matching ambiguity due to noise. Gabor feature/gradient based/edge-aware preserving is shown in Eq. (4).

$$cos\,t_{grad}(p, d) = \left|\nabla_G \mathrm{Img}L(p) - \nabla_G\, Img R\left(p^d\right)\right| \tag{4}$$

where $\nabla_G \mathrm{Img}L(p)$ is the gradient measure in the horizontal direction at pixel p on the input left image. The proposed optimization framework for the WLS model utilized the guidance image *G*, and target image *F* to be filtered formulation of WLS is defined in Eq.(5)

$$E(U) = \sum_{i\in\delta}(U_i - F_i)^2 + \lambda\sum_{i\in\delta}\sum_{j\in N(i)} w_{i,j}(U_i - F_i)^2 \tag{5}$$

Parameter δ coordinate set. λ value stabilized the data and smoothness term. Higher λ value resulted in a more significant smoothing effect on *F*. $N(i)$ denotes the neighborhood of pixel for a square patch size$(2r + 1) \times (2r + 1)$ centered at *r* with pixel coordinate *i*. $w_{i,j}$ is the guidance weight *G* based on different applications.

The above figure (Fig. 2) represents the process flow diagram of the proposed model. The resultant disparity map from SGBM serves as an input to the deep neural network model. The square patches were taken as input for the filter process instead of the 4 connected/8 connected neighborhood approach. The weighted least squares filter modified the neighborhood matrix used in occlusion filling. To solve the issue of smoothness bias in regular deep networks for stereo regression, Stereo Mixture Density Networks (SMD-Nets) were developed.

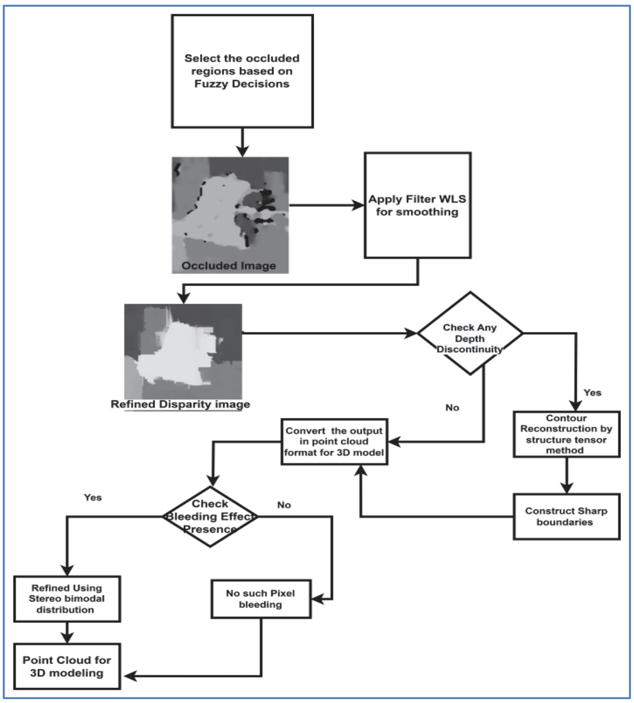
Select the occluded regions based on Fuzzy Decisions
Occluded Image
Apply Filter WLS for smoothing
Refined Disparity image
Check Any Depth Discontinuity
Yes
No
Contour Reconstruction by structure tensor method
Construct Sharp boundaries
Convert the output in point cloud format for 3D model
Check Bleeding Effect Presence
Yes
No
Refined Using Stereo bimodal distribution
No such Pixel bleeding
Point Cloud for 3D modeling

Figure 2 Flow diagram of the contour representation for the images with occluded objects

3.4.1 Fuzzification Applied on Filter

The following table (Tab. 2) denotes the symbol description used in this section.

Table 2 List of Symbols

Symbols	Description
A, $\widehat{A}$	Fuzzy input, output
$W_{nm}(G)$	Guided image weights
a_i, b_i	Linear transformation output of filter coefficients
$\mathrm{st}\left(\nabla g\left(p_0\right)\right)$	Structure tensor for point p_0 invaried area g

Filtering Input image A_n, output image $\widehat{A_n}$ and guidance image G with a filter kernel W_{nm} containing n and m as the pixel indices shown in Eq. (6).

$$\widehat{A_n}=a_iG_n+b_i,\ \forall\in w_i \tag{6}$$

In guided image filtering operation, the output is in the form of a linear transformation with guided image 'G' and window w_i centered with i and their coefficients a_i, b_i given Eq. (7).

$$\widehat{A_n}=a_iG_n+b_i,\ \forall_n\in w_i \tag{7}$$

where $\widehat{A_n}$ is the linear transform of guided image G. This uses a square patch of $2r+1$ and $2r+1$ with radius r. Then the local linear model of $\widehat{A_n}$ has an edge only if an G_n has the edge $\nabla\widehat{A_n}=a\nabla G_n$. To determine the linear coefficients of a_i, b_i we need constraints from the filtering input A. so the output $\widehat{A}$ in Eq. (8) is modeled by subtracting an unwanted pixel due to high-intensity variance in depth inconsistencies n.

$$\widehat{A_n}=A_n-n_i \tag{8}$$

The guided image filtering with ε as the smoothness parameter has decided the degree of smoothness with penalizing large a_i is described in Eq. (9).

$$E\left(a_i,b_i\right)=\sum\nolimits_{n\in w_i}\left(\left(a_iG_n+b_i-A_n\right)\right)^2+\varepsilon a_i^2 \tag{9}$$

$$a_i=\frac{\left(1/|w|\right)\sum\nolimits_{p\in w_i}G_nA_n-\overline{G_i}\,\overline{A_i}}{\sigma_i^2+\varepsilon} \tag{10}$$

The coefficients are valued:

$$b_i=\overline{A_i}-a_i\overline{G_i} \tag{11}$$

where $\overline{G_i}$ and σ_i^2 in Eq. (10) and Eq. (11) are the mean and variance of G_n in w_i. Where $|w|$ denotes the total number of pixels w_i and $\overline{A_i}=\frac{1}{|w|}\sum\nolimits_{p\in w_i}A_n$ is the mean of p in w_i.

So, after computing the coefficients of window w with a and b, the filtering output in Eq. (12).

$$\overline{A_n}=\frac{1}{|w|}\sum\nolimits_{i/n\in w_i}a_iG_n+b_i \tag{12}$$

Noticing that $\sum\nolimits_{i/n\in w_i}a_i=\sum\nolimits_{n\in w_i}a_i$ due to the symmetry of the box window, we have rewritten it as:

$$\overline{A_n}=\overline{a_i}G_i+\overline{b_i} \tag{13}$$

And $\overline{a_i}$, $\overline{b_i}$ in Eq. (13) are average coefficients of all windows overlapping n mentioned in the equations Eq. (14), Eq. (15), and Eq. (16).

$$\overline{A_n}=\left(\frac{1}{|w|}\sum\nolimits_{i\in w_n}a_i\right)G_n+\left(\frac{1}{|w|}\sum\nolimits_{i\in w_n}b_i\right) \tag{14}$$

$$W_{nm}^{GF}\left(G\right)=\frac{1}{|w|^2}\sum\nolimits_{i:(n,m)\in w_i}\left(1+\frac{\left(G_n-\overline{G_i}\right)\left(G_m-\overline{G_i}\right)}{\sigma_i^2+\varepsilon}\right) \tag{15}$$

$$\overline{A_n}=\sum\nolimits_{m\in w_n}W_{nm}^{GF}\left(G\right)A_m \tag{16}$$

Thus the outputof the filtering process was processed with fuzzification parameter ε to decide the patch with the highest partial occlusion chosen for the better resolution with edge-preserving information for effective contour representation.

4 GUIDED FILTER FOR OBJECT CONTOUR REPRESENTATION

The Guided filter for the contour representation of objects was based on the optimal value correspondence with the subject matter and illumination of the image. We hypothesized that Gaussian and Laplacian masks further decreased the errors seen in the border regions. This means that the algorithm can tackle occlusions more effectively because those proposed filters have a smoothing impact in those areas.

4.1 Structure Tensor-Based Weighted Least Square Filter

A well-known edge-preserving technique, which uses the Weighted Least Square filter (WLS) [9], and weighs the highly dependent on image gradients. The smoothing strategy needs to concentrate on areas of local textures with more significant gradients. So, the Structure Tensor-based Weighted Least Square (ST-WLS) [10] method was used to enhance the smoothing effect preserving object boundaries.

Table 3 Gradient measure for different methods

Method	Smoothing Weights, λ	Based on Isotropy	Based on Gradients, σ
Fuzzy + ST-WLS (ours)	640	0.2	1.5
ST-WLS [10]	1 - 1000	0.2	2.0
WLS [9]	8000	63	1.5
Fuzzy Logic [2]	255	27	1

The above table (Tab. 3) describes that the smoothing weight will be smaller if the gradient value is more significant. The final decision made the system to choose the correct patch for the other learning process. It preserved the structure of objects along with the object boundaries for point cloud representation. The smoothing filter WLS takes an input image g, and u the output image obtained is similar to g, once after the smoothing for the whole image is performed instead of focusing on the pixels with high gradients in g, where p is the coordinate/location of a single pixel. The first term reduces the gap between u and g to a minimum. By minimizing the partial derivatives of u, the second term achieved smoothness. The parameter λ affects the relative weight of the two terms. The smoothing weights S_x and S_y in Eq. (17) are spatially variable and are functions of g. If the gradient increases, then S_x and S_y will decrease. Where variable t in Eq. (18) denotes the log-luminance channel of g, (usually between 1.2 and 2) is the exponent that controls the sensitivity measure of gradients g, and ε (generally set to 0.0001) is a small constant aid to avoid division by zero in the stable areas of g, compute S_x and S_y.

$$\sum_p \left((u-g)^2 + \lambda \left(S_x(g) \left(\frac{\partial u}{\partial x} \right)^2 + S_x(g) \left(\frac{\partial u}{\partial y} \right)^2 \right) \right)_p \tag{17}$$

$$S_{y,p} = \left(\left| \frac{\partial t}{\partial y} \right|_p^\alpha + \varepsilon \right)^{-1} \quad S_{x,p} = \left(\left| \frac{\partial t}{\partial x} \right|_p^\alpha + \varepsilon \right)^{-1} \tag{18}$$

A tensor called a structure tensor was used to deduce the underlying structure of a pixel. The pixels around point p_0 fall inside a square window of radius r, which was centered on p_0. Here, we use $(r+1)^2/2$ as our value for 2. Analysis of st eigenvalues reveals the underlying structure of an image. We made λ_1 the smallest eigenvalue of st as 1, and the accompanying eigenvector as v_1 mentioned in Eq. (19), Eq. (20), Eq. (21).

$$st\left(\nabla g\left(p_0\right)\right) = G_\sigma \cdot \left(\nabla g\left(p_0\right) \nabla g\left(p_0\right)^T\right) \tag{19}$$

$$\nabla g\left(p_0\right) = \left[\frac{\partial g \partial g}{\partial x \partial y} \right]_{p_0}^T \tag{20}$$

$$G_\sigma = \frac{1}{z} e^{\frac{-\left\| p - p_0 \right\|_2^2}{2\sigma^2}} \tag{21}$$

The λ_2 the greatest eigenvalue of st, v_2 is also an eigenvector. Therefore we can write it as v_2. The size of an eigenvalue is proportional to the strength of the gradient along its eigenvector. It is challenging for the WLS filter to produce acceptable smoothing results under these conditions. We suggest using a structural tensor-based WLS (ST-WLS) filter. The ST-WLS filter can precisely identify the difference between boundaries with tiny gradients and local textures with big gradients. X denotes the values of the absolute gradient in the horizontal direction, while Y designates the values of the absolute gradient in the vertical direction.

$$w_{x,p} = \left| \left(C \cdot f(X) \right)_p^\alpha + \varepsilon \right|^{-1} \tag{22}$$

$$w_{y,p} = \left| \left(C \cdot f(Y) \right)_p^\alpha + \varepsilon \right|^{-1} \tag{23}$$

$$C = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 \tag{24}$$

where C in Eq. (22), Eq. (23), Eq. (24) denotes the pixel p's isotropy measure. If p is highly isotropic, then C tends to zero when both λ_1 and λ_2 become identical. Then C value becomes close to 1. If p is a strongly anisotropic pixel, then λ_2 is larger than λ_1.

Table 4 Gaussian Kernel size with sigma

S.No.	Kernel Size	Sigma Value
1	13	0.7
2	9	0.5
3	5	0.7
4	3	0.2

Tab. 4 describes how the kernel size was chosen for the Gaussian model along with the normalized value sigma of {0.7, 0.5, 0.2}. Among these, the kernel size of 13 with a sigma value 0.7 has the better convoluting property for the feature extracted using the convolution layer is fed as an input for the fuzzy logic to decide the crisp, fuzzy output representation.

4.2 Occluded Object Detection and Filling

The occlusion detection was carried by comparing occluded areas to non-occluded areas; the accuracy of differences measured in occluded areas is, on average, poorer.

4.2.1 Patch Measure of Occluded Region

The occlusion region occ_e, occ_f in the patch of left image range is $\left[occ_e^1, occ_f^1\right]$, where $occ_e^1 - occ_f^1 < \alpha t$ for some small $\alpha t \simeq 2$. The corresponding region in another image must be wider than $occ_e^1 - occ_f^1$. Consider a partial occlusion region, $\left[x_{1b}, x_{1e}\right]$ in the occlusion map *Ol*, where $x_{1e} - x_{1b} < \eta t$ for few small values of $\eta t \approx 2$. Then, the corresponding region in the other image must be wider than $x_{1e} - x_{1b}$. This can be represented by the disparity $dl\left(x_{1b} - 1\right) > dl\left(x_{1e} + 1\right)$ for *dl* and $dr\left(x_{rb} - 1\right) < dr\left(x_{1e} + 1\right)$ for *r*. Then in a region limit set to $\left[x_{1b}, x_{1e}\right]$ find an actual boundary, assigns different disparity value before and after that boundary.

4.2.2 Handling of Outlier Pixel

One of the most accessible strategies for dealing with an outlier pixel is to set its Disparity equal to the least gap between the disparities of its geographically closest consistent (inlier) pixels on its left and right sides. These pixels may belong to occluded background pixels. In practice, the disparity value was assigned to it and for the nearest consistent (inlier) pixels on the left and right sides.

4.2.3 Hole Filling Process

In this, occluded areas were filled based on the Disparity map-based approach. This filling served best for the sharp contour transition in a 3D model. Compared with the reference or ground truth image, the object contours in the depth map need to be corrected. The image has occlusions, which cause the outlines of the foreground objects to be inconsistent. These were filled using the hole-filling algorithms using WLS (Weighted Least Square) filter along with the interpolation method. The repairment of depth discontinuity has been done based on the interpolation technique embossed in the filter model as described in Algorithm 1.

```
Algorithm 1: For selecting a patch based on fuzzy
Initialize the model parameters.
Train on all the pixels in 1st epoch for input.
For i← from 2 to total epochs, do
Involve pixels of a square patch
Choose the adaptive window across a square patch
Filter occlusion pixel based on the fuzzy decision
Train with/without fuzzy decision
do occlusion filling based on 4-connected neighbor pixels
end for
Obtain the final model
```

4.3 3D Object Reconstruction

The feature-based approach is extracting the particular patch to find the best match for occlusion filling. Then the Gaussian Filter is preferred in this work for effective edge-preserving and smoothing. Typically, noise can be effectively decreased by using a linear filter, such as a Gaussian filter, but the resulting disparity map always has edge fattening.

```
Algorithm 2: Occlusion Detection
for each $I$, when Patch($i$) > 1
    Set $i_m$ = arg max $d(Patch_r(i))$, $i_m = Patch_r(i) - i_m$
    If $i = i_m$, $Visi_i^i\left(i\right) = 1$
    Else $Visi_1\left(i\right) = 0$
end for
```

where for each patch *Patch*(*i*), *I* represent *x* coordinate of left/right images, *d* represents Disparity, $Patch_r(i)$ denotes a set of pixels in the left image. $Visi_1(i)$ denote the visibility function of occlusion for the left image, about a set of pixels in the left image revealing the presence/absence of occluded pixels in the $Patchi_m$ denoting the along m gradient direction for the input image. The pixel-wise reconstruction loss aggregated in different directions using filters is shown in Eq. (25).

$$C_{i,j}^{arg} = \frac{\sum_{x=i-r}^{i+r-1} \sum_{y=j-r}^{j+r-1} w_{x,p} c_{x,p}^{r}}{\sum_{x=i-r}^{i+r-1} \sum_{y=j-r}^{j+r-1} w_{x,p}} \tag{25}$$

The variable $w_{x,p}$ denotes $e^{-\frac{I_{i,j} - I_{x,p}}{2}}$ and intensity pixels of target image are $I_{i,j}$.

5 EXPERIMENTAL SETUP AND RESULTS

The occlusion reasoning is incorporated with two state-of-the-art gradients on the challenging Tsukuba image Pair taken from the Middlebury Stereo [16] Dataset. The Tsukuba and cones dataset can be used as training images for the Middlebury stereo evaluation image dataset, while Venus and Teddy can be used as test image pairs. These conclusions come from the experimental research that was conducted. Middlebury v3 is a relatively small stereo dataset with a high resolution focusing on indoor settings with predetermined lighting conditions.

Table 5 Evaluation Result of Image

Model	Num of 3D Conv Layers	Avg Error Rate (Threshold = 3) / % All Non-*occOcc*			Run-Time / s
ours	6	3.63	3.93	3.46	0.023
Stereo-Net	4	4.83	4.30	4.25	0.072
PSM-Net	8	2.14	4.62	3.86	1.032

Tab. 5 describes different model runtimes with corresponding thresholds. The PSM-Net (Pyramidal Stereo

Matching Network) is time-consuming and uses pyramid spatial pooling and an hourglass network. Stereo-Net employs a mixture density function about the piecewise continuous numbers in which object boundaries may or may not correspond to pixel centers.

Table 6 Different methods of smoothing computational time

Method	WLS	ST-WLS	Fuzzy	Ours-proposed
Time(seconds)	8.02	1.65	2.25	0.32

Tab. 6 describes how much time is taken to process the input image using different methods. The neighborhood construction considers a disparity map along with the outlier's representing occlusion is processed for smoothing with coefficients as $r = 1$, $\lambda = 1$, for the square patch window. Bimodal and Laplacian spaces are employed to model robustly, as the output representation is shown in Eq. (26) and Eq. (27).

$$p(d)=\frac{\pi}{2b_1}e^{\frac{-|d-\mu_1|}{b_1}}+\frac{1-\pi}{2b_2}e^{\frac{-|d-\mu_2|}{b_2}} \tag{26}$$

$$\hat{d}=\underbrace{\arg\max}_{d\in\{\mu_1,\mu_2\}} p(d) \tag{27}$$

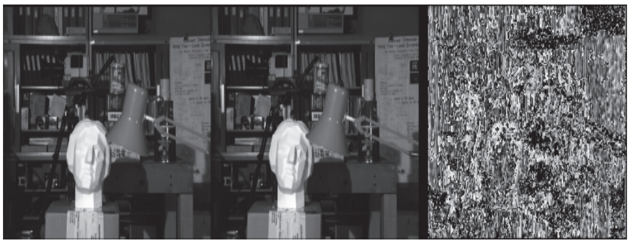

Figure 3 Input images a) left, b) right, c) higher normalization value max = 6400

In Fig. 3a, Fig. 3b and Fig. 3c above, the disparity images for the stereo pair are obtained using the traditional Semi Global Block matching method. They are further processed using adaptive fuzzy systems for handling occluded patch pixels. The normalization of the images has been applied before being processed to identify any outliers present.

In Fig. 4a, Fig. 4b and Fig. 4c the filter we have chosen decided the occlusion region based on the normalization parameter, differentiating pixel intensities. Then the region has undergone smoothing to avoid border discrepancies.

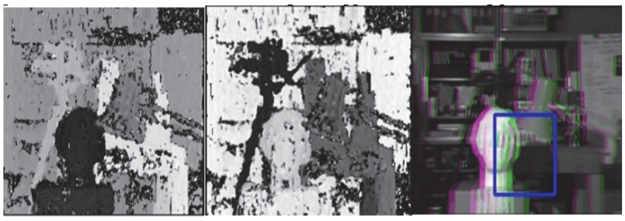

Figure 4 Filter parameter for normalization factor value a) 400, b) 640, c) selective patch for smoothing

Fig. 5a and Fig. 5b describes the fuzzy logic edge detector and the WLS smoothing coefficients based on the logic rule set for the selective patch.

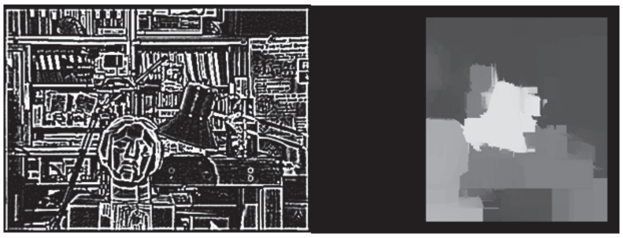

Figure 5 a) fuzzy-based edge detector using LOG, b) smoothened window by WLS method

Fig. 6 denotes the occlusion effects in fuzzy inference system for Fuzzification in selecting patches. Fig. 7a is the conversion output of the depth image into a point cloud for 3D modeling. Fig. 7b is graphical representation of object contours respective to the mean value of the image. The colored bars denote x and y coordinates concerning the theta value as 0.5 for the selection patch received from the fuzzy logic.

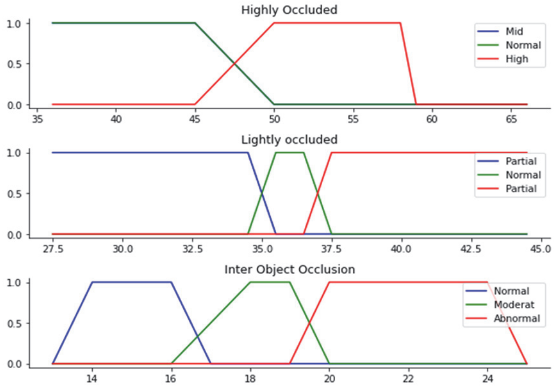

Figure 6 Fuzzy Membership function for Fuzzy Inference System to Fuzzy rules based on Fuzzification

Tab. 7 represents the Disparity, Dimension 1 illustrates the point estimate for the method in 2D convolution, dimension 2 uni-model output representation, and Dimension 5 bimodal representation.

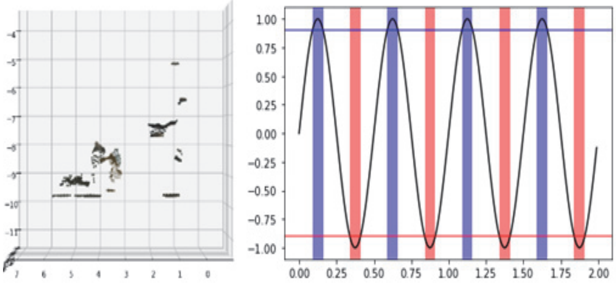

Figure 7 a) Point cloud of input stereo image for 3D reconstruction, b) objects contour under occlusion

EPE metrics (End Point Error), and SEE (Soft Edge Error) calculate the mean (*Avg.*) and standard deviation (σ) metrics for inputs, the fraction of pixels with errors larger than those values are denoted in the table.

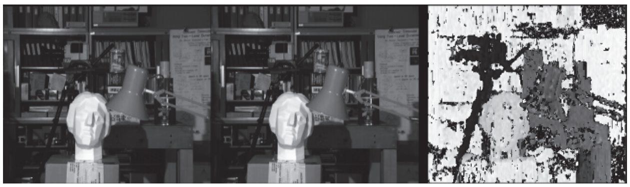

Figure 8 a) left, b) right images, c) smoothening without WLS

Fig. 8 describes how the stereo image disparity maps look without a smoothing effect in the presence of outliers.

Table 7 Evaluation table for different convolutional methods

Met.	Dim.	SEE3			SEE5			EPE	
		Avg.	*σ.* [1]	*σ.* [2]	*Avg.*	*σ.* [1]	*σ.* [2]	*Avg.*	*σ.* [3]
U-Net (2D)	1	1.86	43.67	28.46	1.97	34.65	20.98	1.56	7.18
	2	2.22	44.26	26.72	2.32	38.37	26.57	1.87	10.24
	5	1.65	31.06	13.27	1.25	27.05	16.37	1.18	5.94
PSM (3D)	1	1.68	34.45	19.47	1.45	32.45	18.43	1.1	5.52
	2	2.2	36.8	22.67	2.89	36.19	21.78	1.88	7.43
	5	1.43	24.56	11.54	1.23	23.63	11.52	1.11	4.7
Propos ed	1	3.59	60.16	42.91	3.21	56.49	40.03	4.81	34.6
	2	3.01	62.4	43.29	3.69	59.04	42.19	5.32	41.64
	5	2.23	52.32	32.51	2.61	49.42	32.2	4.05	30.41

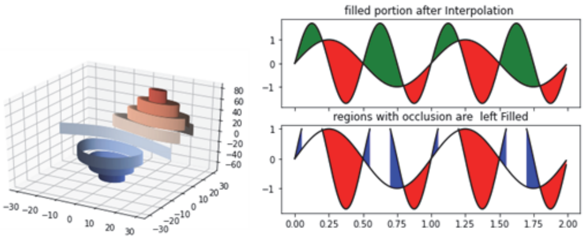

Figure 9 a) contour representation in 3D plot with delta = 0.07, b) filled and unfilled regions of object contour in occlusion

Fig. 9a describes how the contour representation of objects after 3D convolution is performed over fuzzy logic. Then Fig. 9b highlights the regions which are occluded and not occluded with different colors before and after the filling process.

6 CONCLUSION

Based on fuzzy rules, the disparity map was refined using SMD-Nets and the bimodal distribution for effective occlusion handling. Our method employs a continuous function formulation of sharp and precise disparity values at arbitrary spatial resolution. Bimodal mixture densities were used for the output representation to avoid the typical over-smoothing problem in learning-based stereo networks. This finds its application in scene understanding and robot navigation. The experimental results obtained are compared with the existing models. The model's efficiency was improved by selecting a few patches based on the fuzzy approach succeeding in the smoothing process with less memory and time, even in inaccurate regions due to occlusion. In future work, we plan to investigate multi-view stereo images with the adaptive fuzzy method for contour representation.

7 REFERENCES

[1] Hamzah, R. A. & Ibrahim, H. (2016). Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*. https://doi.org/10.1155/2016/874292

[2] Medeiros, M. D., Gonçalves, L. M. G., & Frery, A. C. (2010). Using fuzzy logic to enhance stereo matching in multiresolution images. *Sensors*, *10*(2), 1093-1118. https://doi.org/10.3390/100201093

[3] Kendall, A., Martirosyan, H., Dasgupta, S., Hendry, P., Kennedy, R., & Bachrach, A. (2017). End-to-End Learning of Geometry and Context for Deep Stereo Regression. *2017 IEEE International Conference on Computer Vision (ICCV)*, 66-75. https://doi.org/10.1109/ICCV.2017.17

[4] Liu, R., Deng, Z., Yi, L., Huang, Z., Cao, D., Xu, M., & Jia, R. (2016). Hole-filling Based on Disparity Map and Inpainting for Depth-Image-Based Rendering. *International Journal of Hybrid Information Technology*, *9*(5), 145-164. https://doi.org/10.14257/ijhit.2016.9.5.12

[5] Nieradka, G. (2009). Disparity smoothing algorithm for three-dimensional scene reconstruction. *IFAC proceedings volumes*, *14*(13), 419-424. https://doi.org/10.3182/20090819-3-pl-3002.00073

[6] Zbontar, J. & Lecun, Y. (2019). Stereo Matching by Training a Convolutional Neural to Compare Image Patches. *Computer Vision and Pattern Recognition*, *17*(65), 1-32. https://doi.org/10.48550/arXiv.1510.05970

[7] Zhang, K., Lu, J., & Lafruit, G. (2009). Cross-Based Local Stereo Matching Using Orthogonal. *IEEE Transactions on Circuits and Systems for Video Technology*, *19*(7), 1073-1079. https://doi.org/10.1109/TCSVT.2009.2020478

[8] Hirschmuller, H. (2007). Stereo Processing by Semi-Global Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328-341. https://doi.org/10.1109/TPAMI.2007.1166

[9] Liu, W., Chen, X., Shen, C., Liu, Z., & Yang, J. (2017). Semi-Global Weighted Least Squares in Image Filtering. *2017 IEEE International Conference on Computer Vision (ICCV),* 5861-5869. https://doi.org/10.1109/ICCV.2017.624

[10] Zhang, Z. & Jung, C. (2017). Structure tensor-based WLS filter for adaptive smoothing. *2016 Visual Communications and Image Processing (VCIP),* 23-26. https://doi.org/10.1109/VCIP.2016.7805507

[11] Jang, W. S. & Ho, Y. S. (2014). Efficient depth map generation with occlusion handling for various camera arrays. *Signal, Image Video Process*, *8*(2), 287-297. https://doi.org/10.1007/s11760-013-0550-2

[12] Xie, H., Yao, H., Zhou, S., Zhang, S., Tong, X., & Sun, W. (2021). Toward 3D object reconstruction from stereo images. *Neurocomputing*, 463, 444-453. https://doi.org/10.1016/j.neucom.2021.07.089

[13] Zeljko, S., Igor, B., Mario, S., Milos, K., & Djordje, V. (2022). An Innovative Photogrammetric System for 3D Digitization of Dental Models. *Technical Gazette*, 29(5), 1560-1566. https://doi.org/10.17559/TV-20220401232505

[14] Banno, A. & Ikeuchi, K. (2011). Disparity map refinement and 3D surface smoothing via directed anisotropic diffusion. *Comput. Vis. Image Underst*, *115*(5), 611-619. https://doi.org/10.1016/j.cviu.2010.11.020

[15] Lin, Y., Lu, N., Lou, X., Zou, F., Yao, Y., & Du, Z. (2013). Matching Cost Filtering for Dense Stereo Correspondence. *Mathematical Problems in Engineering*, *2013*(4). https://doi.org/10.1155/2013/654139

[16] Scharstein, D. & Szeliski, R. A. (2002) Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision, 7*(42). https://doi.org/10.1023/A:1014573219977

Contact information:

H. SHALMA
(Correpondng Author)
Department of Computing Technologies, Faculty of Engineering and Technology,
School of Computing, SRM Institute of Science and Technology,
Kattankulathur, Chennai - 603203, Tamil Nadu, India
E-mail: sh3369@srmist.edu.in

P. SELVARAJ
Department of Computing Technologies, Faculty of Engineering and Technology,
School of Computing, SRM Institute of Science and Technology,
Kattankulathur, Chennai - 603203, Tamil Nadu, India
E-mail: selvarap@srmist.edu.in