final project – 2 שיטות מתמטיות לעיבוד לניתוח תמונות

Barak Meiri 308001346

In this document I will explain my final project, which deals with the problem of stereo matching estimation.

The reference paper I based on is:
Stereo Matching with Mumford-Shah Regularization and Occlusion Handling by Rami Ben-Ari and Nir Sochen.

I my project we try to estimate the disparity between two images which are different only by a translation in the horizontal axis.

There are many ways to solve this problem, some use classical methods and some use learning methods to estimate different stages in the process of the disparity estimation.

In this project we focus on the method that is presented in the reference paper, and try to combine optimizations to get the disparity results.

Description and summarization of the reference paper:

This paper introduces a novel approach to solving correspondence in binocular stereo vision. It uses a variational framework with unique regularization terms for discontinuity preservation and occlusion handling. The method provides accurate disparity maps with sharp boundaries and occlusion maps. Experimental tests show significant improvements compared to other methods, making it one of the top-ranked stereo matching algorithms on the Middlebury stereo benchmark.

This paper addresses the problem of correspondence in binocular stereo vision, aiming to establish a disparity map showing pixel differences between two images. It discusses the importance of dense disparity maps for detailed information and categorizes dense matching algorithms into local and global methods. Variational methods, which offer mathematical soundness and continuous solutions, are highlighted as effective for correspondence establishment. The Middlebury stereo benchmark is mentioned as a ranking method for stereo algorithms, and the paper proposes a method that promotes variational approaches in this benchmark. The paper also introduces a data-fidelity term that incorporates image gradients and color information in the objective functional.

This paper addresses the ill-posed nature of the stereo vision minimization problem by introducing a regularization term based on the Mumford-Shah (MS) functional, focusing on piecewise smooth modeling with abrupt disparities. It also evaluates discontinuity maps and introduces additional constraints to improve disparity/depth boundary allocation.

Unlike some previous variational stereo methods, this approach considers half-occlusions in the stereo problem and suggests a novel method for their extraction. The method's performance is assessed on various datasets, including the Middlebury benchmark, showing superior results compared to recent variational methods.

The paper extends previously published work with several notable differences, including a comprehensive introduction to related work, improved discontinuity function calculation, extensive experimental evaluation, and comparisons between L2 and L1 regularizers in the Mumford-Shah functional.

The paper is organized into sections covering related work, the Mumford-Shah framework, the baseline method, an inhomogeneous image-based approach, a novel variational model handling half-occlusions and discontinuities, minimization strategies, and detailed experimental results and comparisons, followed by a conclusion.

<u>Our work in this project includes the following steps:</u>

We start from two stereo images.

Initializing several neural networks following the concepts in the reference paper:

- $\phi(x, y)$ - determines if pixel (x,y) is occlude or not.
- $v(x, y)$ – discontinuity function which determines if (x,y) is a pixel on the contours of the disparity levels.
- d(x,y) – disparity map which tells the disparity for each pixel in the image.

We try to learn these functions (\ mappings) on the pixel space of the two images.

Now we get to the loss and the math behind the method described in the reference paper.

We cite the reference paper equation of which we use to determine the loss:

The Heaviside function is used to binarize the occluded and not occluded regions. $\phi \geq 0$ means visible pixel.

$$H(\phi_l) = \begin{cases} 1, & \phi_l \geq 0 \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

$$H_\epsilon(\phi) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{\phi}{\epsilon}\right)\right)$$

$u(x)$ gives the rational that the disparities f the right image and the left image should be the negative of each other.

$$u_l(\mathbf{x}) := d_l(\mathbf{x}) + d_r(\mathbf{g}_l), \quad \forall \mathbf{x} \in \Omega_l, \tag{15}$$

e_{occl} is the occlusion map that measure the inconsistency in $u(x)$

$$e^l_{occl} = -\ln\left(\epsilon_{occl} + (1 - \epsilon_{occl})e^{-|u_l|}\right). \tag{16}$$

The first loss is the loss on the consistency in $u(x)$:

$$E^l_{occl}(\phi_l) = \int_{\Omega_l} (e^l_{occl} * G_\sigma) H(\phi_l) + t(1 - H(\phi_l)) + \nu \|\nabla H(\phi_l)\| dA, \tag{19}$$

We wish to have small values of $u(x)$ which indicates the consistency in the disparities of the two images.

Now we define the loss on the two images.

Starting with s_d which represent the differences between the images after considering the current disparity between them.

$$s_d^2 = \|I_r(\mathbf{g}_l) - I_l(\mathbf{x})\|^2 + \lambda \|\nabla I_r(\mathbf{g}_l) - \nabla I_l(\mathbf{x})\|^2. \tag{5}$$

If the disparity is correct, we should get small values of s_d.

e_d is a modified L1 version of s_d which is crucial for preventing singular zero.

$$e_d(s_d^2) = \sqrt{s_d^2 + \eta^2}, \tag{4}$$

e_S is a smoothness prior on the images.

$$e_s = \beta \sqrt{\|\nabla d_l\|^2 + \eta^2} \tag{9}$$

This yields the second loss which is:

$$E_L\left(d_l^{\mathbf{v}}, v_l\right) = \int_{\Omega_l} \left[e_d + e_s v_l^2 + \widehat{\alpha}(v_l - 1)^2\right] H(\phi_l) + \widehat{\epsilon}\|\nabla v_l\|^2 dA. \tag{23}$$

This loss takes care of the visible parts, where we can guess the disparity between the two images.

the steps for my algorithm:

Given two stereo images (l, r):

1. initialize NNs: ϕ, v, d
2. calculate e_{occ}
3. calculate E_{occ} loss (loss 1)
4. calculate E_L loss (loss 2)
5. Backward the gradients to the NNs and change their params accordingly.
6. Repeat 2-5 until some condition\convergence.

Experiments:

I used the cones data images:

left image right image

the outputs logs from the algorithm code seems like:

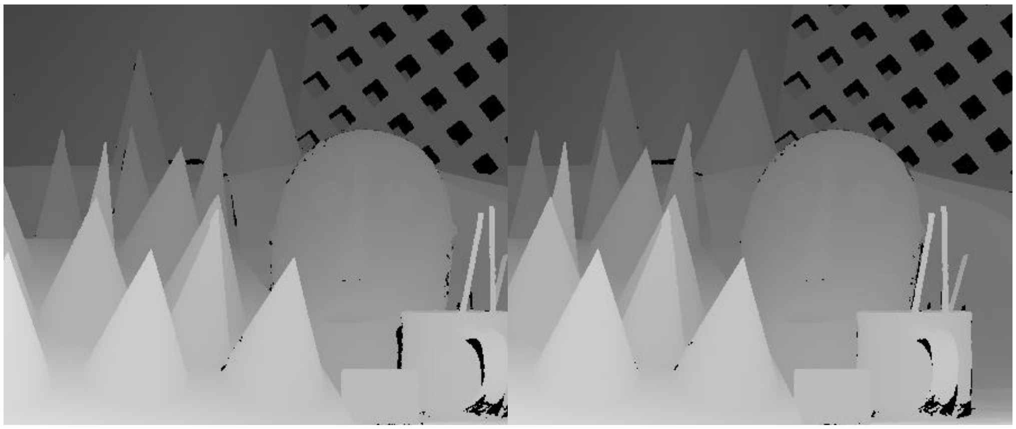
```
Ep: 2   L1-L: 99.35 L1-R: 1.43  L2-L: 570.08    L2-R: 93654.52  es_l: 55.05 es_r: 26.83 ed_l: 102066    ed_r: 94893 phi_l: -6.30    phi_r: 2.95 d_l: 0.98   d_r: 0.99
Ep: 3   L1-L: 99.31 L1-R: 1.32  L2-L: 567.22    L2-R: 93789.14  es_l: 49.37 es_r: 9.89  ed_l: 102066    ed_r: 94928 phi_l: -6.31    phi_r: 3.08 d_l: 0.98   d_r: 1.00
Ep: 4   L1-L: 99.26 L1-R: 1.25  L2-L: 564.07    L2-R: 93860.12  es_l: 44.80 es_r: 9.72  ed_l: 102066    ed_r: 94928 phi_l: -6.33    phi_r: 3.18 d_l: 0.98   d_r: 1.00
Ep: 5   L1-L: 99.22 L1-R: 1.19  L2-L: 560.99    L2-R: 93915.47  es_l: 41.16 es_r: 9.58  ed_l: 102066    ed_r: 94928 phi_l: -6.35    phi_r: 3.27 d_l: 0.98   d_r: 1.00
Ep: 6   L1-L: 99.18 L1-R: 1.14  L2-L: 558.01    L2-R: 93960.65  es_l: 38.31 es_r: 9.45  ed_l: 102066    ed_r: 94928 phi_l: -6.37    phi_r: 3.36 d_l: 0.98   d_r: 1.00
Ep: 7   L1-L: 99.14 L1-R: 1.10  L2-L: 555.12    L2-R: 93998.70  es_l: 36.25 es_r: 9.34  ed_l: 102066    ed_r: 94928 phi_l: -6.39    phi_r: 3.44 d_l: 0.98   d_r: 1.00
Ep: 8   L1-L: 99.09 L1-R: 1.06  L2-L: 552.32    L2-R: 94031.50  es_l: 34.73 es_r: 9.24  ed_l: 102066    ed_r: 94928 phi_l: -6.41    phi_r: 3.51 d_l: 0.98   d_r: 1.00
Ep: 9   L1-L: 99.05 L1-R: 1.03  L2-L: 549.61    L2-R: 94060.20  es_l: 33.46 es_r: 9.15  ed_l: 102066    ed_r: 94928 phi_l: -6.43    phi_r: 3.58 d_l: 0.98   d_r: 1.00
Ep: 10  L1-L: 99.01 L1-R: 1.00  L2-L: 546.98    L2-R: 94085.66  es_l: 32.32 es_r: 9.07  ed_l: 102066    ed_r: 94928 phi_l: -6.45    phi_r: 3.65 d_l: 0.98   d_r: 1.00
Ep: 11  L1-L: 98.96 L1-R: 0.98  L2-L: 544.44    L2-R: 94108.46  es_l: 31.32 es_r: 9.00  ed_l: 102066    ed_r: 94928 phi_l: -6.47    phi_r: 3.71 d_l: 0.98   d_r: 1.00
```

Which tells the first and second losses values and the intermediate values of different meaningful metrics.

First of all I tried to estimate the disparity in a single side. Later when I saw that it didn't work, I did the symmetric calculations also for the other side.

In order to do a sanity check, I tried also to insert the ground truth disparity to check that the losses have smaller values then without inserting the ground truth – and it did happen.

The disparity for the sanity test was:



Left disparity   right disparity

As we can see also the GT is not perfect since there are occluded regions.

To summarize, my project deals with the stereo matching problem and I aim to solve it with optimizations over the Mum-Shah losses and the loss for the occlusions.