Robust Depth Estimation with Occlusion Detection Using Concepts of Optical Flow

Hiral Raveshiya
Computer Engineering Department
Vidya Vikas Education Trust's Universal College Of Engineering
Thane,India.
hiral.raveshiya@universal.edu.in

Ankita Sanghavi
Computer Engineering Department
Sal Institute of Technology and Engineering Research
Ahmedabad,India
ankita.gandhi@sal.edu.in

***Abstract*— In this paper we present an approach to go beyond the accuracy limits of current optical flow estimators. We have used coarse to over-fine approach along with hybrid interpolation to upgrade coarse to fine approach. Coarse to fine approach is used by most modern optical flow algorithms. Results of suggested method show benefit for sub-pixel motion. It also reduces the estimation error to great extent. Hybrid interpolation method is used which is an integration of bilinear and bi-cubic interpolation methods. Results of our approach on benchmark sequences show that estimated depth map are clearer and boundaries are sharper than original coarse to fine approach with bi-cubic interpolation method. Once optical flow vectors are found occlusion is detected based on the concept of residual error image.**

Keywords—Coarse to over-fine optical flow; Hybrid interpolation; Occlusion; Optical Flow; Residual error image

I. INTRODUCTION

Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene. Optical flow techniques such as motion detection, object segmentation, time-to-collision and focus of expansion calculations, motion compensated encoding, and stereo disparity measurement utilize this motion of the objects' surfaces and edges. Optical flow estimation is still one of the key problems in computer vision. Estimating the displacement field between two images, it is applied as soon as correspondences between pixels are needed. In the recent times the quality of optical flow estimation methods has increased dramatically. Starting from the original approaches of Horn and Schunck [1] as well as Lucas and Kanade [2], research developed many new concepts for dealing with shortcomings of previous models. In order to handle discontinuities in the flow field smoothness constraints was introduced that permit piecewise smooth results [3, 4]. Coarse-to-fine strategies [5, 6] as well as non-linearised models [7, 8] have been used to tackle large displacements. In this paper we suggest extending the multi-scale approach beyond the fine grid to over-fine levels. The representation of the image data on the over-fine grids is obtained by interpolation. Coarse to over-fine approach leads to sharpening of the flow edges. Sub-pixel motion is observed to be more accurately estimated as well. Occlusion phenomena are a critical component of the image construction process, shaping the statistics of natural images. Occlusions arise when a portion of the scene is visible in one image, but not another. In Da Vinci Steropsis, portions of the scene that are visible from the left eye are not visible from the right eye, and vice-versa. Typically occlusion occurs at depth discontinuities in video stream. Once optical flow is estimated occlusion detection can be performed in many different ways. In our work we have used concept of residual error image.

The rest of this paper is organized as follows: Section II introduces variational optical flow method along with model assumptions. Section III discusses coarse to fine strategy with warping that is heart of many modern optical flow estimation algorithms. The details of coarse to over-fine approach is given in section IV. Section V introduces hybrid interpolation method. Concepts of occlusion and occlusion detection are explained in section VI. The experimental results are presented in section VII followed by brief summary. This paper is extension of [17] which presents our work of optical flow estimation.

II. VARIATIONAL OPTIC FLOW METHOD

Calculus of variations is a field of mathematics that deals with functionals, as opposed to ordinary calculus which deals with functions. Such functionals in our case are formed as integrals involving an unknown function and its derivatives. The interest is in extreme functions: those making the functional attain a maximum or minimum value. Nowadays, variational methods are among the best performing techniques in image processing for depth-map reconstruction: being global methods and thus operating on entire image domain, they recover the depth-map as the minimizer of a suitable functional, which we will call energy functional. First let us discuss constraints used in our method. We have used model assumptions suggested by Brox [9].

A. Grey value constancy assumption

Since the beginning of optical flow estimation, it has been assumed that the grey value of a pixel is not changed by the displacement.

$$I(x, y, t) = I(x + u, y + v, t + 1) \qquad (1)$$

Here I denote a rectangular image sequence and w = (u, v, 1) is the searched displacement vector between an image at time t and another image at time t + 1. In (1) we assume that grey value remains constant in both images, I(x,y) at time t and I(x+u,y+v) at time t+1. I(x+u,y+v) is the image after displacement added to the image I(x,y).

B. Gradient constancy assumption

The grey value constancy assumption has one decisive drawback: It is quite susceptible to slight changes in brightness, which often appear in natural scenes. Therefore, it is useful to allow some small variations in the grey value and help to determine the displacement vector by a criterion that is invariant under grey value changes. Such a criterion is the gradient of the image.

$$GI(x, y, t) = GI(x + u, y + v, t + 1) \qquad (2)$$

Where G denotes the spatial gradient. In (2) we assume that gradient remains constant in both images, I(x,y) at time t and I(x+u,y+v) at time t+1. I(x+u,y+v) is the image after displacement added to the image I(x,y).

C. Smoothness Assumption

The smoothness term stands for the assumption that the neighboring regions belong to the same object and thus these regions have similar depth. The main role of the smoothness term is the redistribution of the computed information and smoothing of depth outliers. In case we get no reliable information from the data term, the smoothness term will realize its smoothing effect by filling in the problem region with data, calculated from neighboring regions. In fact, we introduce here an additional assumption that the depth-map is globally smooth – a smoothness assumption.

Let us suppose that we are given a stereo image, represented as a number of pictures of a certain scene from different positions and angles Ii (u, v), i=1…N, where (u,v) denotes the point coordinates in image. And moreover let us suppose that we would like ψ to compute the depth-map field z(u, v). According to variational methods we should construct an energy functional which has the following structure:

$$E(z(u,v)) = \iint_{\Omega} F\left(I_1(u,v), \ldots, I_N(u,v), z(u,v), \frac{\partial z(u,v)}{\partial u}, \frac{\partial z(u,v)}{\partial v}\right) dudv \qquad (3)$$

The functional E (z (u,v)) consists of two terms: the data term and the smoothness term. While the data term provides us with the information about depth, the smoothness term distributes this information:

$$E(z) = \iint_{\Omega} \text{Data\_term}(u,v,z) + \text{Smoothness\_term}(Zu,Zv) dudv \qquad (4)$$

Such an approach guaranties us that the computed depth-map will be always dense. For example, in a case of data term cannot give us in some region more or less useful information, the smoothness term fills in this region with information, computed from neighboring regions.

Our main goal is to find such a function z (u,v) , which minimizes the energy functional E(z(u,v)). By another words, having constructed the energy functional, we should minimize it in order to find the best solution for the depth- map. The main energy functional minimized for the computation of optical flow is a combination of appearance and smoothness terms of the images involved. Optical flow typically tries to compute a flow field that minimizes the difference between the two images involved. The main energy functional can be expressed as

$$E(u,v) = E_{Data}(u,v) + \alpha\, E_{smooth}(u,v) \qquad (5)$$

$$E_{Data}(u,v) = \int_{\Omega} \psi(|I(x+w) - I(x)|^2 + \gamma|\Delta I(x+w) - \Delta I(x)|^2)\, dx \qquad (6)$$

$$E_{smooth}(u,v) = = \int_{\Omega} \psi\, (|\Delta_3 u|^2 + |\Delta_3 v|^2)\, dx \qquad (7)$$

The idea in [3] was to minimize the above functional globally (for all the pixels simultaneously). The data term is a number of combined assumptions that certain features of the image do not change, but remain constant from one picture to another. The data term is responsible for supplying us with information about depth-field and stand for the constancy assumptions that are used. Here data term represents gradient constancy assumption. And smoothness term stands for the assumption that the neighboring regions belong to the same object and thus these regions have similar depth. Here α represents the regularization term. Once data term and smoothness terms are constructed energy functional is derived and then we can use euler-langarnge equation. For improvement we additionally introduce a ψ function, which is in general a penalizing function.

III. COARSE TO FINE STRATEGY WITH WARPING

The coarse-to-fine strategy follows two main aims. The first aim is to solve the problem of the multiple minima in the energy functional and the problem of avoidance of local minima during the iteration process. And the second aim is to tackle the problem of large displacements. For the purpose of the implementation technique the literature offers us two different strategies: the scale-space focusing method that considers the problem at different smoothness scales, keeping the picture's resolution unchanged; and the multiresolution technique that considers the problem at different resolution levels [13]. In this paper we will use the multiresolution technique since it is much more efficient from the computational point of view.

In the figure 1 we can see an example of a coarse levels pyramid, built from a single image and its downscaled instances. At the top of this pyramid the coarsest level is situated and at the bottom – the fine level, the original picture. If we have N coarse levels, then the fine level will be the first coarse level and the coarsest one will be the N-th coarse level. The depth-map reconstruction process becomes with this technique an iteration process. We reconstruct a depth-map on a coarse level image and then use this reconstructed depth-map as initial map for the next coarse

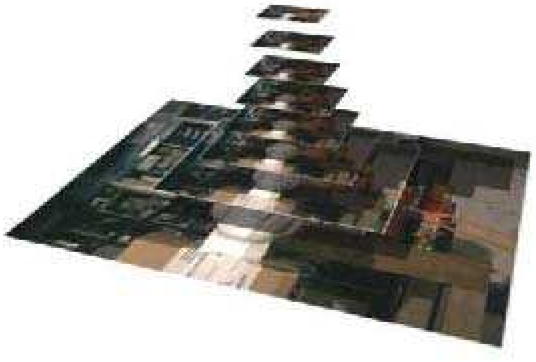

Fig 1. Multiresolution coarse levels pyramid 2D example

level.When we have the original picture, to build the pyramid, illustrated in figure 2, it is enough to choose the number of coarse levels and the dimensions of the picture in the coarsest level. It is very important to get the first depth- map from the coarsest level, because it will be a fundament for all the following computations. That's why the best choice of the dimensions of picture at the coarsest level is such choice where the largest displacement in optic flow will be sufficiently small. The number of layers N should be chosen in such a way, that the width increment be smaller or equal to the width of the picture at the coarsest level.

It will guarantee that the displacement from coarse level to coarse level will not increase more than in two pixels. The coarse-to-file levels technique starts with the variational process on a coarse instance of the original picture, where the displacement much smaller (less than 4 pixels). Coming from a coarser level to a finer level, we have the initial depth-map, calculated from the previous step, and thus we just step by step increase the displacement and recalculate the depth-map with better accuracy. Warping denotes the distortion of the image sequence which is required for the compensation for the already computed motion. Theoretical justification of the warping technique is provided in [9].

IV. COARSE TO OVER-FINE STRATEGY

Consider a multi-scale optical flow estimation algorithm A, whose objective functional E (A) is comprised of a data term, Ed, and a smoothness term, Es. The following modified algorithm [11] A0 yields a significantly more accurate flow than A:

1) Let I0(x, y, t) denote the input sequence, and let k = 1 denotes the iteration counter.

2) Execute A on the original image sequence I0.

3) For each of the input frames, extend the image pyramid beyond the finest scale, using interpolation (resulting in a coarse-to-over fine pyramid). The extension of the pyramid is an additional frame sequence, Ik, derived from the original sequence, Ik-1, in the following manner:

$$Ik(x, y, t) = Ik\text{-}1(x/2, y/2, t) \qquad (8)$$

For non-integer indices x/2, y/2 of Ik-1, either bilinear or bicubic interpolation is used.

4) Continue the execution of A on the extended pyramid.

5) Sample the resulting flow back to the original grid.

6) Increment k and return to step 3, repeat procedure for predefined number of iterations.

The coarse to over-fine approach presented here is an effective way to improve the accuracy of optical flow estimation algorithms. Coarse to over-fine approach estimates the edges in the frames very clearly that is useful for detecting boundaries of the different objects of the frames easily. It also provides higher accuracy for smooth sub-pixel motion.

V. HYBRID INTERPOLATION

This method generates high-resolution images with merits of both bilinear and bicubic interpolation method. Steps are as follows [12]:

1) Interpolate a low resolution image f0 by applying bilinear interpolators and obtain the linear interpolated result f1.

2) Interpolate image f0 by applying bi-cubic method and obtain nonlinear results f2.

3) For these two interpolated results, we endow the hybrid parameters ρ and (1- ρ) respectively, where (0≤ ρ≤1). And then hybrid result f3 is obtained by summing up these two weighted results.

Thus, hybrid method can be proposed and given by

$$I = \rho A + (1-\rho) B, (0 \leq \rho \leq 1) \qquad (9)$$

Where I, A and B are hybrid interpolator, linear interpolator and bi-cubic interpolator respectively. ρ is hybrid parameter. Hybrid interpolation method is very simple and very easy to implement. The only thing to be considered is the proper selection of the hybrid parameter. It considers both high and low frequency components of the image and thus it are a method with the high accuracy. This concept can be added for interpolation at intermediate stages of coarse to fine strategy with warping in optical flow estimation techniques. Instead of using bi-cubic or biliniear method, hybrid interpolation gives better output. In our experiments we have used 0.312 as value of hybrid parameter.

VI. OCCLUSION

Occlusion means that there is something you want to see, but can't due to some property of your sensor setup, or some event. Exactly how it manifests itself or how you deal with the problem will vary due to the problem at hand.

Let us see some examples to understand the occlusion concept: If we are developing a system which tracks objects (people, cars, ...) then occlusion occurs if an object we are tracking is hidden (occluded) by another object. Like two persons walking past each other, or a car that drives under a bridge. The problem in this case is what you do when an object disappears and reappears again.

If we are using a range camera, then occlusion is areas where you do not have any information. Some laser range cameras works by transmitting a laser beam onto the surface we are examining and then having a camera setup which identifies the point of impact of that laser in the resulting image. That gives the 3D-coordinates of that point. However, since the camera and laser is not necessarily aligned there can be points on the examined surface which the camera can see but the laser cannot hit (occlusion). The problem here is more a matter of sensor setup.

The same can occur in stereo imaging if there are parts of the scene which are only seen by one of the two cameras. No range data can obviously be collected from these points. In optical flow estimation we are having frames which are taken at different time instances, say t and t+1. Here occlusion can be termed as the portion of the frame which is visible in the frame taken at time t, but the same portion is not visible in the frame taken at time t+1(or reverse). Applying optical flow algorithm on two frames we can get the distance vectors for each pixel of the frame. Based on this these vectors we can find occluded regions between two frames.

We have detected occlusion between two frames using following two steps [14]:

1) Find residual error image between two frames

2) Find the high residual areas from residual error image using some threshold value Residual error image- it is the image which contains the difference of two images: one original image and another one is generated from the original image.

We have found residual error image from two frames: original frame (Any one frame from two input frames) and Warped frame, warping is performed on other frame which is not used in the first step above.

For finding residual error image, we have warped any one of the frame. For warping we are using velocity vectors u and v, which is explained in section 6.2. For example if we are warping the frame at time t+1 then we will find the difference between original frame at time t and the warped frame at time t+1. Exactly reverse case is equally correct.

Basically, residual error image contains the difference between two images in terms of the pixel. Once residual error image is obtained, comparison against some threshold value is performed. Threshold value here is in terms of number of pixels like 3 pixels or 5 pixels or some other value.

Consider that we are using 5 pixels as threshold value. In simple words, we have to check residual error image and find the areas from it which contains the value greater than 5 pixels.

There is simple logic behind this method which is explained here in clear words. Here warping of the image is performed based on the u and v, it means the vectors which shows the displacement of every pixel of the frame. So when we perform warping using u and v, we are getting frame at time t+1 in which each pixel has moved by certain pixel position. And then we are finding the pixels which have moved more than some desirable level that is threshold level. This is obvious thing because if the pixels has moved more than allowable level and we are getting much more difference between two frames then it may be the case that pixels in frame at time t has moved so far at time t+1 and they have become occluded.

VII. EXPERIMENTAL RESULTS

Following outputs show the result of occlusion detection experiments. Ground truth motion is used to compute residual error image. Using residual error occlusion maps are generated. Following figures show the result of occlusion detection experiments and compares my results with ground truth results.

Army sequence

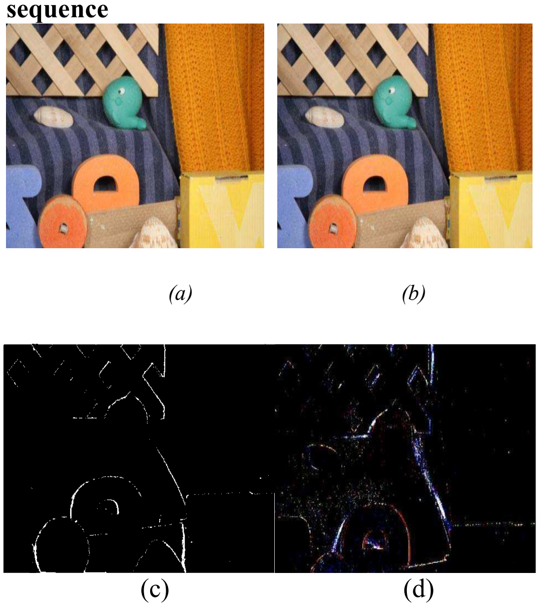

Fig 2. Output comparison (a) input fame-army frame 10 (b) input frame-army frame 11 (c) ground truth occlusion (d) occlusion detected my code

Hydrangea Sequence

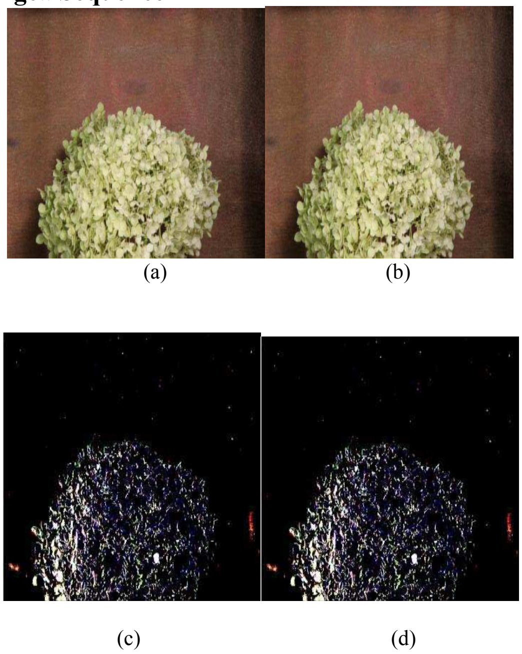

Fig 3. Output comparision (a) input fame-hydrangea frame 10 (b) input frame-hydrangea frame 11 (c) ground truth occlusion (d) occlusion detected my code

Our algorithm combines the concepts of coarse to over-fine, hybrid interpolation and occlusion detection. Experiments done for various image sequences but following table shows analysis for army frame sequence 7 & 8. Reduced error rate shows that proposed method is robust under noisy conditions. Results are not much affected by changing the inner and outer iterations. AAE is the average angular errors which gives the optimal result. Compared to original algorithm computation time is also reduced.

TABLE I. Comparision Table

Reduction Factor	**Outer Iterations**	**Inner Iterations**	**SOR Iterations**	**Computation Time**	**AAE**
0.95	3	100	10	5 min	1.94
0.90	3	50	10	2.5 min	2.05
0.85	5	100	10	8 min	2.54
0.80	1	100	10	1.3 mn	3.40

VIII. FUTURE WORK AND CONCLUSION

Coarse to fine approach is widely used in current optical flow estimators; it enables estimation of optical flow at small areas of frames. Augmenting coarse to fine approach with coarse to over-fine approach improves accuracy of depth map. Coarse to over-fine gives better output based on concepts of over sampling. We have performed hybrid interpolation on input frame sequence which improves the output of depth map. Based on values of calculated flow vectors residual error image can be calculated which is useful in detecting occlusion. Residual error calculated in our algorithm is very accurate that provides us results near to ground truth occlusion maps. Computation time and AAE is also reduced in proposed algorithm.

The generated depth map can be used to generate 3D representation of the image from the given sequences of frames taken at different times. Occlusion filling can be performed after occlusion detection. Using Joint Bilateral Filter image resolution can be increased for sharpen object boundaries and less errors.

REFERENCES

[1] Master's Thesis Multi – View 3D Reconstruction with Variational Method by Sergey Kosov

[2] http://en.wikipedia.org/wiki/Optical_flow

[3] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[4] Variational Optic Flow Computation: From Continuous Models to Algorithms by Joachim Weickert, Andr´es Bruhn, Nils Papenberg, and Thomas Brox,ECCV 2005, pp 35-40.

[5] T. Brox, A. Bruhn, N. Papenberg, J. Weickert. High accuracy optic flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, Computer Vision ECCV 2004, volume 3024 of Lecture Notes in Computer Science, pp 25-36. Springer, Berlin, 2004.

[6] Secrets of Optical Flow Estimation and Their Principles Deqing Sun Brown University Stefan Roth TU Darmstadt Michael J. Black Brown University. In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, California, June 2010.

[7] Optic Flow Goes Stereo: A Variational Method for Estimating Discontinuity-Preserving Dense Disparity Maps by Natalia Slesareva, Andr´es Bruhn, and JoachimWeickert, DAGM-symposium 2005:33-40.

[8] Coarse to Over-Fine Optical Flow Estimation by Tomer Amiaz , Eyal Lubetzkyy Nahum Kiryati , Pattern Recognition40(9):2496-2503(2007)

[9] An iterative hybrid image interpolation method by yan tian, caifang zhang,fuyuan peng and sheng zheng, In proceedings of ICIC(1)' 2005. Pp. 10~19

[10] http://vision.middlebury.edu/flow/data

[11] A Variational Method for Scene Flow Estimation from Stereo Sequences by huguet, F and Devernavy,F. ICCV-2007. IEEE 11th international conference.

[12] Occlusion Detection and Motion Estimation with Convex Optimizationby Alper Ayvaci, Michalis Raptis ,Stefano Soatto, Advances in Neural Information Processing Systems; 2010.

[13] Thesis-Robust Incremental Optical Flow, Michael Julian Black ,YALEU/CSD/RR #923 September 1992

[14] Robot Maximilian. http://www.howtoandroid.com/.

[15] I. A. Ypsilos, A. Hilton, S. Rowe. Video-rate capture of dynamic face shape and appearance. In Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 117-122. May 2004.

[16] F. Kücükay, J. Bergholz. Driver assistant systems. Scientific report. Could be found on http://www.osd.org.tr/

[17] Robust Depth Estimation Using Concepts of Optical Flow by Hiral Raveshiya and Viral Borisagar. In proceedings of ETCSIT-2012. Pp 147~151.