Cyclopean optical flow

Anonymous ECCV submission

Paper ID 1234

Abstract. This paper presents... (70 a 150 mots).

Keywords: Optical flow, Stereo, ...

1 Introduction

This paper is concerned with the optical flow problem, and more specifically 1D optical, suitable to also be called, in the presence of a suitable epipolar geometry, small baseline stereo reconstruction.

2 Previous work

Optical flow estimation is a problematic approached by a variety of techniques, dominated by latest developments in deep learning and a strong background on traditional models (energy-based, gradient-based, block-matching, feature-tracking, etc.). (((Cite)))

As mentioned in [8], artificial neural networks start to outperform existing models according to public performance over data-sets such as [1] [5] (((cite more))). (((make a little explanation of each dataset)))

(((explain limitations of excisting models)))

Some of these outstanding results have been attained by [9] [11] (((cite more and cite related to regression))).

(((Make little explanation of each deep learning article)))

Despite the latter progress, deep learning encounters limitations that motivate further research in the domain of optical flow (((Mention the limitations))). Non-learning algorithms show relevance due to their robustness, independence from training, (((mention more and cite a source))) and potential collaboration with other motion estimation techniques to enhance a solution (((cite some examples))).

It is read in (((cite))) the multiple applications optical flow can have for example: object segmentation, ego motion, (((cite more))) etc.(((Mention small baseline stereo)))

2.1 small baseline stereo reconstruction

Stereo vision is an important domain of computer vision which is focused on constructing the 3D feature of depth. The main characteristic of stereo is the use of two cameras displayed horizontally. In account of the latter, the two pair of images are projected in the same geometric plane (co-planar) and have a geometric disparity between them. In case images are not co-planar, they can be rectified into a shared plane. Some methods to re-project such images are: planar, cylindrical and polar rectification with epipolar geometry.

In [2][6] it's said that a big angle between views is generally expected to reach good precision, in contrast, [2] small baseline stereo is proposed. The latter assumed that the disparity between two images vary slowly in space and reinforces that small angles (which generate fewer occlusions and a more similar image) favor the estimates of depth.

2.2 Optical flow

Optical Flow is the perceptible 2D motion of a scene through time. As optical flow is mostly a visual experience, to interpret displacement, information needs to be extracted from images to identify distribution similarities and then determine their variance in position.

The information that is commonly employed is the intensity of an image, which is the value of it's pixels.

Horn and Schunk [7], introduced brightness consistency and small motion assumptions into optical flow, which enables pixel to pixel comparison while maintaining a system with linear behavior. This method is also constrained by a 2D projection as representation of a 3D physical world, making estimation more accessible without highly compromising accuracy during the solution's approximation.

But, as mentioned, [7]'s method is an approximation limited by it's own initial hypothesis, although it's results are remarkable, it welcomes to development in order to increase performance and robustness.

(((metion lucas-kanade, etc.)))

3 Cyclopean optical flow

3.1 Image interpolation

An image is a common visual representation given by pixels and their distribution but this representation might become deficient when computing optical flow due to derivative computation; accordingly, derivatives are notoriously sensitive to noise and can be inaccurate when computed over discrete values, being images a natural source of these defects (((cite something))). To tackle this problem, image interpolation is proposed to transform pixels into a representation which features continuity in it's values and first derivative.

Likewise, the cyclopean method is highly dependant on intensity interpolation due to the lack of control over the start point of a displacement vector. This will be further discussed on subsection 3.2.

From now on, I will refer to the interpolation of any discrete image $\ddot{I}$.

An interpolation of first order yields discontinuous first derivatives, which introduces difficulties on computation. Discontinuity results in a low resolution derivative, which can translate in rough changes along spatial exploration which won't allow an accurate perception of the image's shape. Also, the low capacity to represent the gradient of an image will have a higher probability to land in a constrained zone where it's solution will be undefined, further discussion in subsection 3.4.

Therefore, the interpolation form applied is a cubic spline interpolation [10] where $I(x, y) = \ddot{I}[x, y]$ to respect cardinality constraints. The degree of this spline grants enough capacity to ensure continuity on the first and second derivatives. To avoid instability at the function's edges, padding is recommended to ensure influence of the original representation still precedes.

3.2 Cyclopean Optical Flow

1D Cyclopean optical flow is a technique to estimate the one dimensional cyclopean displacement in an image's intensity when it's scene is affected by motion. In a sequence of 2 images, the first element (I_1) defines the pixel's original state, the following segment (I_2) will acquire I_1's altered distribution due to the existing dynamics. The theoretical average intensity that would be observed in the middle of the transition from I_1 towards I_2 is defined as "cyclopean image" (C). Using C's intensity values as a reference allows to decompose the cyclopean solution into two parts: the flow from I_1 to C and the flow from C to I_2. Both velocities are then added up and are assumed to represent the movement behavior between I_1 and I_2.

Considering that the intensity will be interpolated, gradient-based estimation methods [3] is of big interest to exploit the result of approximating discrete values to a continuous function.

In a continuous representation of C, I_1 and I_2, the cyclopean values are computed as follows:

$$C(x) = \frac{I_1(x) + I_2(x)}{2} \tag{1}$$

The accuracy of this model is expected to increase due the inclusion of every image derivatives instead of a single one. Including both ($I_1'(x)$ and $I_2'(x)$) the method traces more information making it possible to identify unusual behavior and avoid finding fake solutions. In subsection 3.4 this issue will be addressed in greater depth. This model is also enhanced due to the computation of two smaller motions instead of a large one. The introduction of two displacements artificially increases in the sample rate, allowing to capture more details on the scene's transition offering a smoother solution. Figure (((put figure))) illustrates the traditional and cyclopean methods and the properties mentioned are graphically identifiable.

This is also mathematically perceived through their optical flow definitions. Traditional optical flow is computed as:

$$v = \frac{I_1(x) - I_2(x)}{I_1'(x)} \tag{2}$$

The equation 2 is derived from a Taylor series approximation [3], it only considers the spatial derivative of I_1 to estimate the displacement v. On the other hand, cyclopean optical flow is computed as:

$$\begin{aligned} v &= v_1 + v_2 \\ v_1 &= \frac{I_1(x) - C(x)}{I_1'(x)} \\ v_2 &= \frac{C(x) - I_2(x)}{I_2'(x)} \end{aligned} \tag{3}$$

Having two estimations influencing v's value, constrains the result offering a more reliable result.

It's important to mention that the cyclopean solution of point p is equivalent to the solution of $p - v_1$, therefore, all motion vectors will be represented with an initial start point of $p - v_1$.

3.3 Iterative estimation

According to the traditional method, optical flow can exactly be computed when the starting point and the destination point have the same gradient [3]. In (((ref of figure 1))) The spatial derivative acts as the slope of a straight line which will be constrained by $p - v^{n+1}$ and $p - v^n$ in x, due to the euclidean behavior of the system, v is found immediately.

In the cyclopean method, we use a reference point in between the intensity functions $I_1(x)$ and $I_2(x)$ for each point p referred to as: cyclopean value c_p computed as (1). The latter is a theoretical value in between $I_1(p)$ and $I_2(p)$, which will allow to decompose $v = v_1 + v_2$ where v_1 is the cyclopean solution of pixel p towards the function I_1, accordingly for v_2 and I_2. If the gradient of I_1 at points p and $p - v_1$ are equal, v_1 is found immediately due to the euclidean behavior. This concept is also applied to I_2 at points at p and $p + v_2$ to estimate v_2; nevertheless, this gradient exigence is mostly the attribute of a basic gray scale and it does not appears often in practice. In more common systems where the gradient changes spatially, the solution would be just a rough approximate, making it more challenging to find a good solution.

To increase accuracy in non linear cases, it's proposed to estimate v_1 and v_2 with an iterative approach. Starting from an initial guess, optical flow will be updated along with new states of gradient, until convergence.

Equation (4) is used to update and approximate v_1 through iteration, with v_1^n representing v_1 at step n, and an initial value of $v_1^0 =$ (((Any real number?))). This represents the search of c_p's intensity in $I_1(x)$. The process will become redundant once $p - v_1^n$ satisfies $I_1(p - v_1^n) = c_p$, where p's theoretical intensity has been successfully tracked and found in $I_1(x)$.

$$v_1^n = v_1^{n-1} + \frac{I_1(p - v_1^{n-1}) - c_p}{I_1'(p - v_1^{n-1})} \tag{4}$$

The same computation is performed for v_2, with the update function (5).

$$v_2^n = v_2^{n-1} + \frac{I_2(p + v_2^{n-1}) - C(p)}{I_2'(p + v_2^{n-1})} \tag{5}$$

As presented above, using an iterative approach can solve optical flow when the gradient is changing spatially. However, there are more cases to consider in order to compute optical flow robustly, discussed in subsection 3.4.

3.4 Gradient Constraints

It's reasonable to expect that at convergence $I_1'(p - v_1) = I_2'(p + v_2)$, meaning that I_1 has a cyclopean displacement of $v = v_1 + v_2$ at point p in an ideal way. Being the gradients equal at $p - v_1$ and $p + v_2$, they share sign and it's assumed that gradients will keep the same sign throughout the process of iteration. If it happens otherwise, it's considered that the cyclopean optical flow cannot be estimated at p, since it probably won't converge.

Case 1 ("Sign") is illustrated as fig. (((reference to figure))) and then defined:

$$v^n = \begin{cases} v^n = v_1^n + v_2^n, & \text{if } I_1'(p - v_1^n) * I_2'(p + v_2^n) > 0 \\ undefined, & \text{otherwise} \end{cases}$$

Also, the magnitude of gradient is a denominator in equation 4, therefore, a small gradient would result on a big estimate of motion impacting negatively on the computation. Low gradient is to be considered as a lack of texture in images which isn't enough information to compute optical flow. A threshold k will be introduced to define the lowest value of gradient the method can accept in order to have a good approximation.

This behavior will be considered as Case 2 ("Magnitude"), illustrated as fig. (((reference figure))) and defined as:

$$v^n = \begin{cases} v^n = v_1^n + v_2^n, & \text{if } I'(p) > k \\ undefined, & \text{otherwise} \end{cases}$$

If we enforce all previous gradient constraints (cases of gradient sign and magnitude), it is clear that some points will remain unsolved. In order to estimate optical flow for all points, we will then need to "fill the gaps" by using other methods, introduced in subsection 3.5.

3.5 Extra Methods

Scale of I will be modified as an attempt to find good results. The motivations behind this decision, is that a simpler form of I will be more stable due to it's smoother shape. Also, the construction of multiple images at various scale levels combine the information of a larger spatial neighborhood. We proposed to use a pyramidal representation I^n where n is the present level of resolution.

Reducing the scale of an image has been used often in computer vision to alter information with the goal of obtaining a preferable version of the image's signal [3][4]. A pyramid is a multilevel image representation which computes n number of sub-samplings of the reference.

It's fairly common to divide size by a factor of 2 each time reduction is applied, therefore, in the algorithm proposed a down-sampling value of 2 is fixed.

To avoid loss of information, blur effect is applied to load individual pixels of adjacent information. The usual method for blurring computes the equivalent of a pixel with a weighted window. The end result of moving the window over every pixel is a smooth version of the input signal.

With fitting values of weights, the process of smoothing will filter high frequencies signals and sudden changes of intensity. Thanks to this low pass filter behavior, the correlation between pixels is highlighted and noise discontinuities that might deform the scene are restrained.

Gaussian blur uses a Gaussian function to define the weight values to smoothen a signal. The Gaussian behavior will manage importance in a radial form, being the center the most significant, which means the result will pay a lot of respect to the pixel's original form. This trait is important to preserve the essence of the signal's shape after each reduction, that's why it's proposed to use a Gaussian Pyramid for multilevel representations.

3.6 Solving optical flow over multiple scales

The iterative solving equations (4) and (5) will be referred to as the function (6) where v_0 is the initial value of v.

$$v = upgrade(v_0) \tag{6}$$

This definition only considers the base level of an image's pyramid. To solve optical flow over multiple scales, the same upgrade function is applied updating I^{n-1} after solving v^n for I^n.

The iteration starts at the smallest scale (top of the pyramid) and is upgraded as follows:

$$\begin{gathered} v^n = upgrade(v_0) \\ v^{n-1} = upgrade(v^n) \\ \dots \\ v^0 = upgrade(v^1) \end{gathered} \tag{7}$$

At each level, if the gradient does not satisfy the gradient constraints, we stop and keep the latest acceptable solution.

The base level of an interpolated step edge generally won't meet the gradient constraints, but it'll have a wider reach when explored in different resolutions.

The base level of an image's pyramid usually contains more details and noise compared to it's higher levels, these characteristics are not ideal for optical flow estimation due to the discontinuities it produces in the spatial derivatives; however, during the process of sub-sampling, signals will smoothly be altered until no longer precise. This limitations motivate the algorithm to iterate over every pyramid level from lowest resolution I^n to I^0, with the objective of estimating v^0 eluding the gradient constraints by referencing previous solutions of v.

During the process of using the update function in 1, a convergence map can be visualized if the input values of v are used as a two component argument. In (((figure with convergence plot of only one initial update))) is identifiable that convergence also depends on the possible combinations of v_1 and v_2 to form v for a given p. This characteristic is then used to increase the robustness of the method, by relying on the capacity of v to be arbitrary. v_0 will be now a list of initial values $listV_0$ as an input for 6, resulting in a list of possible solutions $listV$ for a given point p.

4 Small baseline Stereo

Link between stereo et optical flow.

4.1 epipolar geometry

Here we describe what happens when small baseline stereo is done... The epipolar line, if they can be assumed to be parallel, will be aligned to maximze parallax.

...

5 Experiments and Discussion

5.1 Optical flow

5.2 Stereo results

Table 1. Font sizes of headings. Table captions should always be positioned *above* the tables. The final sentence of a table caption should end without a full stop

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes ...**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Headings.** Text follows ...	10 point, bold
4th-level heading	*Remark.* Text follows ...	10 point, italic

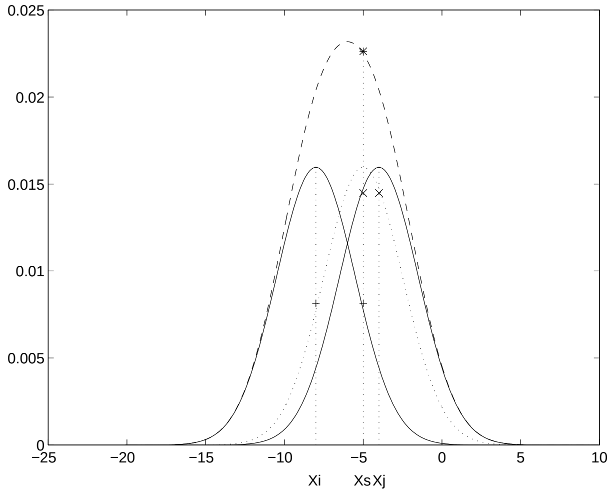

Fig. 1. One kernel at x_s (*dotted kernel*) or two kernels at x_i and x_j (*left and right*) lead to the same summed estimate at x_s. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption. The last sentence of a figure caption should generally end without a full stop

6 Limitations

7 Conclusions

The paper ends with a conclusion.

References

1. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) Computer Vision – ECCV 2012. pp. 611–625. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
2. Delon, J., Rougé, B.: Small Baseline Stereovision. Journal of Mathematical Imaging and Vision **28**(3), 209–223 (Oct 2007). https://doi.org/10.1007/s10851-007-0001-1, https://hal.archives-ouvertes.fr/hal-02269064
3. Fleet, D.J., Weiss, Y.: Optical flow estimation (2005)
4. Fortun, D., Bouthemy, P., Kervrann, C.: Optical flow modeling and computation: A survey. Computer Vision and Image Understanding **134**, 1–21 (2015). https://doi.org/https://doi.org/10.1016/j.cviu.2015.02.008, https://www.sciencedirect.com/science/article/pii/S1077314215000429, image Understanding for Real-world Distributed Video Networks
5. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361 (2012). https://doi.org/10.1109/CVPR.2012.6248074

6. Herath, D.C., Kodagoda, K.R.S., Dissanayake, G.: Modeling errors in small baseline stereo for slam. In: 2006 9th International Conference on Control, Automation, Robotics and Vision. pp. 1–6 (2006). https://doi.org/10.1109/ICARCV.2006.345368
7. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial Intelligence **17**(1), 185–203 (1981). https://doi.org/https://doi.org/10.1016/0004-3702(81)90024-2, https://www.sciencedirect.com/science/article/pii/0004370281900242
8. Hur, J., Roth, S.: Optical flow estimation in the deep learning age (2020)
9. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks (2016)
10. McKinley, S., Levine, M.: Cubic spline interpolation. College of the Redwoods **45**(1), 1049–1060 (1998)
11. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume (2018)