Motion Estimation Using Optical Flow Concepts

Hiral Raveshiya
Department of Computer Engg.,
Government Engineering College, Sector 28,
Gandhinagar, India
hiral_er2005@yahoo.co.in

Viral Borisagar
Asst. Professor
Department of Computer Engg.,
Government Engineering College, Sector 28,
Gandhinagar, India
viralborisagar@yahoo.com

Abstract—motion estimation is demanding field among researchers. The most general and challenging version of motion estimation is to compute an independent estimate of motion at each pixel, which is generally known as optical or optic flow. In our paper we have provided overview of some basic concepts behind optical flow and motion estimation. The paper is intended as starting point for beginners. Some common techniques and application of motion estimation is provided for the review of readers who are interested in this field.

I. INTRODUCTION

Motion is very important feature of image sequences. Motion reveals the dynamics of scenes by relating spatial image features to temporal changes. Motion estimation is a challenging and fundamental problem of computer vision. Optical flow is the only accessible parameter from sequences of 2-D images. There are varieties of applications which uses estimated optical flow field as input for subsequent processing steps. Example applications include motion detection, motion compensation, motion-based data compression, 3-D scene reconstruction, autonomous navigation and the analysis of dynamical processes in scientific applications.

There are many difficulties in motion estimation mainly because of the inherent differences between the optical flow and the real motion field. Horn [1] has suggested possibility of 3-D reconstruction from motion fields. Though the motion field itself is often inaccessible, the motion field can almost always be unambiguously related to translational and rotational velocities of rigid surfaces. Real-world sequences include many problems like transparent overlay of multiple motions, occlusions, illumination changes, nonrigid motion, stop-and-shoot motion, low signal-to-noise (SNR) levels, aperture problem and correspondence problem and many more. According to Verri and Poggio[2] - the true motion field is hardly ever accessible and suggests that only qualitative properties of the motion field should be computed[3]. Due to these problems motion estimation is difficult but these problems are usually not spread over entire image area and that makes motion analysis practicable in many applications. These problems also pose a constraint on optical flow computation: errors have to be detected and quantified. Error measurement is highly important in quantitative scientific measurement applications. While more qualitative requirements of standard computer vision applications require precise and dense optical flow fields in order to reduce the spreading of errors into subsequent processing steps. Despite all of these problems, continuous research is going on in motion estimation field. There has been a latest stimulation of interest in low-level motion estimation in the literature [4, 5, 6, 7]. Barron et al. [6] provides overview of existing optical flow techniques. Recent trend of research is towards quantitative performance analysis of optical flow techniques and with this trend motion analysis is also gaining increasing interest. Due to advancement in computer hardware, in terms of both speed and memory, more and more complex and computationally expensive techniques have become applicable in recent years. Before some years ago computer vision applications were restricted to two consecutive images. Increased power of hardware extends computer vision into the temporal domain that led to new techniques with increasing performance.

The remaining paper is organized as follows: Section II explains some fundamentals of optical flow estimation. Section III compares optical flow techniques and correspondence techniques and outlines the appearance of motion in image sequences. Optical flow based motion estimation techniques are reviewed in section IV. Section V introduces concept of multi frame motion estimation followed by conclusion.

II. OPTICAL FLOW

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. In
simple words: 2D Motion Field can be defined as 2D velocities for all visible points. And Optical Flow Field can be defined as Estimate of the 2D motion field. 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 [8] as well as Lucas and Kanade [9], research developed many new concepts for dealing with shortcomings of previous models. To determine optical flow we need to track some property of images. Two key problems in optical flow estimation are: 1) Determine what image property to track 2) Determine how to track it. Some features of the images are assumed to stay constant among multiple frames during optical flow estimation. Generally used constancy assumptions are discussed below.

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.

\[ I(x, y, t) = I(x + u, y + v, t + 1) \]  \hspace{1cm} (1)

Here I denotes 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 \( t + 1 \).

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 grey value, which can also be assumed not to vary due to the displacement.

\[ G(x, y, t) = G(x + u, y + v, t + 1) \]  \hspace{1cm} (2)

Where \( G \) denotes the spatial gradient.

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.

Computation of optical flow means computation of two vectors- \( U \) and \( V \). \( U \) represents horizontal velocity of motion and \( V \) represents vertical velocity of motion. Usually \( U \) and \( V \) are computed using the concepts of energy functional. And the main aim is to minimize this energy functional. Energy functional consists of two terms: data term and smoothness term. There are number of algorithms available for formulation of energy functional and its minimization. In this paper we are not going into the details of algorithms.

III. FLOW AND CORRESPONDENCE

Optical flow based techniques and correspondence-based techniques are subject of discussion among developers. These both techniques have their own advantages and disadvantages. In this paper we will recall which method seems to be best suited under certain circumstances. In order to solve the aperture problem a variety of optical flow-based approaches have been proposed that try to minimize an objective function pooling constraints over a small finite area. They conclude that differential techniques, such as the local weighted least squares method proposed by Lucas and Kanade [10], perform best in terms of efficiency and accuracy. Phase-based methods [11] show slightly better accuracy but are less efficient in implementation and lack a single useful confidence measure. Analysis of various motion estimation techniques is shown in Jähne [12] and [13]. According to this analysis the 3-D structure tensor technique yields the best results with respect to systematic errors and noise sensitivity. On the other hand, correspondence-based techniques are less sensitive to illumination changes. They are also capable of estimating long-range displacements of distinct features that violate the temporal sampling theorem. In this case any optical flow-based technique will fail. However, correlation-based approaches are extremely sensitive to periodic structures. With nearly periodic inputs they tend to find multiple local minima [5]. Comparative studies show that correlation-based techniques produce unpredictable output for straight edges, while optical flow based techniques correctly estimate normal flow. Correlation techniques also perform less effectively in estimating subpixel displacements than do optical flow-based techniques [5, 14]. Especially at very small displacements in the order of less than
10 pixel/frame, optical flow-based techniques yield better results. Before we turn towards a description of the various techniques, we want to draw the conclusion that neither correlation nor optical flow-based techniques are perfect choices in any case. If the temporal sampling theorem can be assured to be fulfilled, optical flow based techniques are generally the better choice. In other cases, when large displacements of small structures are expected, correlation-based approaches usually perform better. For both kinds of techniques, it is important to get confidence measures in addition to the optical flow. No technique is without errors in any case. Only if errors can be detected and quantified can the result be reliably interpreted. It also shows that differences in precision, attributable to details of the initial formulation, are, in fact, a result of different minimization procedures and a careful numerical discretization of the used filters.

IV. OPTICAL FLOW BASED MOTION ESTIMATION TECHNIQUES

In this section we will focus on common optical flow-based techniques [22].

A. Differential techniques

Local weighted least squares: Assuming the optical flow \( f \) to be constant within a small neighborhood \( U \), Lucas and Kanade [10] propose a local weighted least squares estimation technique. The estimated optical flow is given by the solution of the following minimization problem:

\[
f = \arg \min_{|e|} \left\| |e| \Delta f + \hat{g} \right\|^2 \text{d}x\]

with a weighting function \( w(x) \) selecting the size of the neighborhood. In practical implementations the weighting is realized by a Gaussian smoothing kernel. Additionally, \( w \) could weight each pixel according to some kind of confidence measure, for example, the magnitude of the gradient. In that way, a priori known errors are not propagated into the optical flow computation.

Second-order techniques: Instead of grouping constraints over a local neighborhood, it has been proposed to use second-order information to solve for both components of \( f \) [15,16,17]. Thus, second-order techniques, just as first-order techniques, do not allow for estimating the 2-D optical flow field \( f \) in case of an aperture problem within a local neighborhood. In order to obtain second-order differential information, first-order properties of the image area need to be related over an increased area compared to first-order differentiation. From first-order information the full 2-D optical flow can only be extracted if the spatial orientation changes within the region of interest. Bainbridge-Smith and Lane [9] conclude that first-order differential techniques, such as proposed by Lucas and Kanade [10], are in fact generalized second-order techniques, because they implicitly require variation of the gradient within the region of support.

Global constraints: Local least squares techniques minimize the brightness change constraint over a localized aperture, defined by the size of the spatial window function. Global constraint methods extend the integration to the entire image area and combine the local gradient constraint with a spatial coherence assumption. The resulting objective function consists of two terms. The first one contains the local data (brightness) conservation constraint at each pixel and a second one contains the spatial relation between optical flow vectors:

\[
||e_i||^2 = ||e_i||^2 + \lambda \sum_j (\nabla \theta_j)^T f + \hat{g}_j \right\|^2 \text{d}x + \lambda \sum_j ||e_i||^2 \]

All approaches incorporating global constraints have in common that they result in systems of differential equations relating spatial variations of the optical flow within the entire domain \( D \). Such a system can only be solved iteratively using numerical iteration methods, such as Gauss-Seidel iteration or successive overrelaxation. Although efficient iterative solutions have been developed in numerical mathematics, they are still slower than closed solutions. Another problem of iterative solutions is the question of convergence, which may depend on image content.

B. Tensor-based techniques

In order to determine local orientation Bigün and Granlund [18] proposed a tensor representation of the local image brightness distribution. Starting with a different idea, Kass and Witkin [19] came to a solution that turned out to be equivalent to the tensor method. Searching for a general description of local orientation in multidimensional images, Knutsson [20, 21] concluded that local structure in an \( n \)-dimensional domain can be represented by a symmetric \( n \times n \) tensor of second order. In the analysis of data with a dimensionality higher than two it turns out that using scalars and vectors is no longer always convenient. Tensors-a generalization of the vector concept-are perfectly suited to describe symmetries within local neighborhoods in multidimensional spatial and spatiotemporal signals.
Tensor-based technique is very useful in optical flow estimation but as this paper is basically for the beginners reference we would not provide tensor based technique in much detail.

C. Multifeature techniques
Multifeature (or multiconstraint) techniques try to use two or more features to obtain overconstrained equation systems at the same location. These features have to be linearly independent in order to solve for both components of \( f \). Otherwise the aperture problem remains, leading to singularities in the overconstrained system of equations. Multiple features can be obtained by:

- using multiple light sources and/or multispectral cameras
- visualizing independent physical properties of the same object; and
- using results of (nonlinear) functions of the image brightness.

Of course, all features have to move with the same velocity. Otherwise the estimated optical flow exhibits the motion of the combined feature vector rather than the real object motion. This prerequisite can be violated for features showing different physical properties that are subject to dispersion.

V. MULTI-FRAME MOTION ESTIMATION

So far, we have looked at motion estimation as a two-frame problem, where the goal is to compute a motion field that aligns pixels from one image with those in another. In practice, motion estimation is usually applied to video, where a whole sequence of frames is available to perform this task. One classic approach to multi-frame motion is to filter the spatio-temporal volume using oriented or steerable filters in a manner analogous to oriented edge detection. Because the pixel motion is mostly horizontal, the slopes of individual (textured) pixel tracks, which correspond to their horizontal velocities, can clearly be seen. Spatio-temporal filtering uses a 3D volume around each pixel to determine the best orientation in space–time, which corresponds directly to a pixel’s velocity. Unfortunately, in order to obtain reasonably accurate velocity estimates everywhere in an image, spatio-temporal filters have moderately large extents, which severely degrades the quality of their estimates near motion discontinuities. An alternative to full spatio-temporal filtering is to estimate more local spatio-temporal derivatives and use them inside a global optimization framework to fill in textureless regions. Another alternative is to simultaneously estimate multiple motion estimates, while also optionally reasoning about occlusion relationships. Motion estimation can be cast as a global energy minimization problem that simultaneously minimizes brightness compatibility and flow compatibility terms between key frames and other frames, in addition to using robust smoothness terms. The multi-view framework is potentially even more appropriate for rigid scene motion, where the unknowns at each pixel are disparities and occlusion relationships can be determined directly from pixel depths. However, it may also be applicable to general motion, with the addition of models for object accelerations and occlusion relationships. Let us consider some of the applications of multi frame motion estimation [15].

A. Video denoising
Video denoising is the process of removing noise and other artifacts such as scratches from film and video. Unlike single image denoising, where the only information available is in the current picture, video denoisers can average or borrow information from adjacent frames. However, in order to do these without introducing blur or jitter (irregular motion), they need accurate per-pixel motion estimates.

B. De-interlancing
Another commonly used application of per-pixel motion estimation is video de-interlacing, which is the process of converting a video taken with alternating fields of even and odd lines to a non-interlaced signal that contains both fields in each frame. Two simple de-interlacing techniques are bob, which copies the line above or below the missing line from the same field, and weave, which copies the corresponding line from the field before or after. The names come from the visual artifacts generated by these two simple techniques: bob introduces an up-and-down bobbing motion along strong horizontal lines; weave can lead to a “zippering” effect along horizontally translating edges. Replacing these copy operators with averages can help but does not completely remove these artifacts. A wide variety of improved techniques have been developed for this process, which is often embedded in specialized DSP chips found inside video digitization boards in computers. A large class of these techniques estimates local per-pixel motions and interpolates the missing data from the information available in spatially and temporally adjacent fields.

VI. CONCLUSION

In this paper we have provided overview of some fundamental concepts related to motion estimation using optical flow estimation techniques. Though the paper does not provide details of mathematical concepts of motion estimation, one can have little inside of concepts. Motion estimation techniques are
attractive research area which is described here briefly. There is vast scope of development in multi frame motion estimation also. Depth estimation using motion/optic flow estimation is also growing field, it has many interesting applications like robotics, driver assistant systems, medical imaging and too many to list here.

REFERENCES

Richard Szeliski September 3, 2010 draft