Generalized optical flow in the scale space

Haifeng Gong *, Chunhong Pan, Qing Yang, Hanqing Lu, Songde Ma

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China

Abstract

Scale space is a natural way to handle multi-scale problems. Yang and Ma have considered the correspondence between scales, and proposed optical flow in the scale space. In this paper, we generalize Yang and Ma’s work to generic images. We first generalize the Horn–Schunck algorithm to multi-dimensional multi-channel image sequence. Since the global smoothness constraint for regularization is no longer suitable in general cases, we introduce localized smoothness regularization. In scale space optical flow, points in original image trends to aggregate at a large scale, so we introduce aggregation density as an additional smoothness coefficient. At last, we apply the proposed methods to color images and 3D images.

1. Introduction

Scale space is a natural way to handle multi-scale problems. The correspondence between scales leads to tracking points in the scale space. Optical flow algorithm is used to accomplish the tracking. This idea was originally proposed by Yang and Ma [20]. In this paper, we will generalize the original scale-space optical flow (SSOF) to multi-dimensional multi-channel image filtering.

Scale plays an important role in biological vision and is very important in human vision learning in early life [4]. The scale concept and the notion of multi-scale representation are of crucial importance in signal processing and computer vision. Since the initial work by Thurston and Rosenfeld [16], Klinger [7], this concept has been greatly developed and a series of methods have been proposed, e.g., pyramid, wavelet, and scale space theory.

Optical flow is very important in biological vision too. It is a basic mechanism for visual translation, rotation, and expansion perception [19]. Psychophysical research illustrated the importance of optical flow for the control of posture and locomotion and for the perception of self-motion [19]. Scale-space optical flow is something like perceiving visual expansion with optical flow [14] in biological vision, this is the bionic basis of our method.

In order to extend scale-space pullback to color images, we first generalize the Horn–Schunck algorithm to multi-dimensional multi-channel image sequence. The regularization in multi-dimensional multi-channel Horn–Schunck is more complicated than single-channel 2D version. When the number of channel is greater than or equal to the dimensionality, the intensity constraint equation tends to be over-determined, therefore, it is necessary to determine whether regularization is needed or not. So we introduce localized smoothness regularization into multi-dimensional multichannel Horn–Schunck. In scale-space optical flow, as the scale increases, image points in the original image trend to aggregate into several clusters and the point around the cluster centers are more important. We use the flow density to emphasis the intensity constraint of dense points in image filtering, which results in another type of localized smoothness constraint. At last, we apply these methods in color image and 3D image scale-space pullback. Contribution of this
paper lies in two aspects: a multi-dimensional multi-channel version of Horn–Schunck optical flow and the color image scale-space pullback.

This paper is organized as follows. Section 2 summarizes previous work on scale-space pullback. Section 3 describes generalized optical flow in the scale space. Sections 4 and 5 apply the proposed technique to RGB color image and 3D image, respectively. The conclusions are included in Section 6.

2. Previous work

Scale space is a framework for early vision, which was first proposed by Iijima [18] in 1959, and became popular later on by the works of computer vision researchers [9, 17]. They showed that the natural way to represent an image at finite resolution is by convolving it with Gaussian at different bandwidth, thus obtaining a sequence of blurred image at different scales. It is therefore possible to trace the evolution of certain image structures, such as critical points, over scale. The exploitation of various scales simultaneously has been referred to as deep structure by Koenderink [8].

A problem in linear scale space theory is that image blurring using Gaussian kernels distorts features in the original image. The distortion becomes intolerable at large scales. In order to overcome this difficulty, anisotropic diffusion has been intensively studied [11, 13]. Unfortunately, the effect of anisotropic diffusion is still unsatisfactory in many cases. A much more thorough method is to track points in the scale space. Although this idea has been reported in the literature, it is seldom performed to obtain a better multi-scale representation. It is often believed that this is due to computational complexity. However, Yang and Ma [20] argued that the main reason is that the tracking problem is ill-posed, and a procedure of regularization must be introduced.

Tracking paths in the scale space can be viewed as optical flow in the scale space. In $E(x, y, t)$’s multi-scale representation $\tilde{E}(x, y, t)$, the scale $t$ can also be regarded as “time”. The optical flow $(u(x, y, t), v(x, y, t))$ of $E(x, y, t)$ can be carried out step by step. Thus we obtain a mapping from the original image $E(x, y, 0) := E(x, y)$ to the blurred image $\tilde{E}(x, y, t)$

$$\phi_t(x, y) := (\phi^x_t(x, y), \phi^y_t(x, y)).$$

That is, for fixed $(x, y)$, $\{\phi_t(x, y)|t \geq 0\}$ is an integral curve of the vector field $(u, v, 1)$ with initial condition $\phi_0(x, y) = (x, y)$. Then we are able to define a new image $\tilde{E}(x, y, t) := E(\phi^x_t(x, y), \phi^y_t(x, y), t)$.

In general, the mapping $\phi_t$ is not one-to-one because there may exist singular points in $(u(x, y, t), v(x, y, t))$. However, $\tilde{E}(x, y, t)$ is uniquely determined by $E(x, y, t)$. In terms of differential geometry [5], $\tilde{E}(x, y, t)$ is the pullback function of $E(x, y, t)$ by $\phi_t$. The tracking process can be regarded as a simulation of visual expanding from near to far away from the object of interest. Fig. 1 demonstrates this technique.

The concept tracking back was often used in the literature. However, this is an inaccurate concept because the tracking path in previous tracking strategy is not well-defined. Yang and Ma’s method overcomes the difficulty. To avoid possible confusions, they introduced standard mathematical concepts, and called $\tilde{E}(x, y, t)$ the intrinsic multi-scale representation of $E(x, y)$ at scale $t$, which removes detail structure and preserves only salient structures without heavy distortion. It can be used as a preprocessing for salient structure extraction or segmentation.

$E(x, y, t)$ and $\tilde{E}(x, y, t)$ are related by a coordinate transformation, but their perceptual effects are quite different. $\tilde{E}(x, y, t)$ is much more natural than $E(x, y, t)$. Since regularization is introduced to compute the optical flow, the tracking is well-defined and robust.

Yang and Ma’s work is limited to gray scale images only. In this paper, we will extend their work to the most general cases. To achieve this goal, two problems need to be solved: (1) the classic Horn–Schunck algorithm must be generalized to multi-channel images; (2) since the global smoothness constraint for regularization is no longer suitable in general cases, the localized constraints must be considered.

3. General optical flow

In order to extend scale-space pullback to color images, a multi-dimensional multi-channel version of Horn–Schunck algorithm with local smoothness constraint will be given in this section.

Optical flow is widely used in motion estimation and image alignment [15]. During the past two decades, many methods for the estimation of optical flow have been proposed [1]. According to Arredondo et al. [1], these
algorithms can be classified into three groups: differential techniques, region-based matching, and frequency-based methods. See [3] for comparison. Below, we will introduce a multi-dimensional multi-channel version of Horn–Schunck algorithm.

Let’s give a brief introduction about the classical Horn–Schunck optical flow. Let \( E(x, y, t) \) be the gray level image sequence, if the intensity of patch on an object remains constant, we can get the intensity constraint equation \( E_x u + E_y v + E_t = 0 \). The problem can be converted to the minimization of the energy functional by smoothness regularization.

\[
J[u, v] = \int \int \left( ||E_x u + E_y v + E_t||^2 + x^2 (||\nabla u||)^2 + ||\nabla v||^2 \right) \, dx \, dy. \tag{3}
\]

Its iterative solution is

\[
u^{n+1} = \frac{E_x u^n + E_y v^n + E_t}{\alpha^2 + E_x^2 + E_y^2},
\]

\[v^{n+1} = \frac{E_x u^n + E_y v^n + E_t}{\alpha^2 + E_x^2 + E_y^2}. \tag{5}\]

In practice, we always come in front with multi-channel images. For example, an RGB image, or a more general case, multi-spectral image or a color image plus some texture components. And image lattice can be more than two-dimensional, such as MRI image. A more general case is \( n \)-dimensional image with \( m \) color channels. Color image optical flow has been studied by former researchers, but as far as we known, no work on general multi-dimensional multi-channel image optical flow has been reported. A color image optical flow algorithm was proposed by Ohta [12], in this algorithm smoothness constraint was removed, and intensity constraint equation of each channel was used to solve the optical flow vector. Arredondo et al. computed optical flow using textures [1], they first estimated the optical flow in intensity and textural images independently, and then combine these estimates by weighting them according to the gradient in each channel. Channel weighted optical flow can be seen in [2], but it applied weights over channels, and can be considered as an equivalent of color space transformation. Here, we will give the Horn–Schunck algorithm for multi-dimensional multi-channel image in this section.

Let \( E = E(X, t) = (e_1(X, t), e_2(X, t), \ldots, e_n(X, t))^T \) be the \( n \)-dimensional image sequence with \( m \) color channels, where \( X = (x_1, x_2, \ldots, x_n)^T \) is an \( n \)-dimensional vector. Our task is to compute the optical flow vector between two consecutive frame \( E(X, t) \) and \( E(X, t+\Delta t) \) in the sequence.

In continuous form, the iso-intensity constraint is \( \frac{dE(X, t)}{dt} = 0 \), and the intensity constraint is

\[
E_x V + E_t = 0, \tag{6}
\]

where \( V = (v_1, v_2, \ldots, v_n)^T \) is the optical flow velocity vector to be solved. We follow the notational convention that the partial derivatives with respect to a column vector are laid out as a row vector. Partial derivatives of field \( E(X, t) \) with respect to \( X \) and \( t \) in (6) are defined as:

\[
E_x = \frac{\partial E}{\partial X} = \begin{pmatrix}
\frac{\partial e_1}{\partial x_1} & \cdots & \frac{\partial e_1}{\partial x_m} \\
\vdots & \ddots & \vdots \\
\frac{\partial e_n}{\partial x_1} & \cdots & \frac{\partial e_n}{\partial x_m}
\end{pmatrix},
\]

\[
E_t = \begin{pmatrix}
\frac{\partial e_1}{\partial t} \\
\vdots \\
\frac{\partial e_n}{\partial t}
\end{pmatrix}^T. \tag{7}
\]

Our problem is to solve the intensity constraint equation (6). Define the error for image intensity as \( E_{b}^2 = ||E_x V + E_t||^2 = (E_x V + E_t)^T (E_x V + E_t) \), and the smoothness constraint as \( E_{c}^2 = \sum_{i,j} (\frac{\partial^2}{\partial t^2})^2 \). The energy functional to be minimized is

\[
J[V] = \int (E_x V + E_t)^T (E_x V + E_t) + x^2 \sum_{i,j} \left( \frac{\partial^2}{\partial x_i \partial x_j} \right)^2 \, dx. \tag{9}
\]

The Euler–Lagrange equation is

\[
E_x E_x V + E_y E_t - x^2 \nabla^2 V = 0, \tag{10}
\]

where \( \nabla^2 V = (\nabla^2 v_1, \ldots, \nabla^2 v_n) \) is Laplacian of the optical flow velocity vector.

Let \( y_i = (y_{i1}, \ldots, y_{in})^T, i = 1, \ldots, N \), be a set of grid points in image lattice, \( N \) is the number of pixels or samples, and \( u_i = V[y_i] \) be the optical flow at sample points. Laplacian of \( V \) can be approximated by Laplacian of Gaussian:

\[
\nabla^2 V = (\nabla^2 v_1, \ldots, \nabla^2 v_n) = \sum_i u_i \text{LoG}(X - y_i, \sigma) \tag{11}
\]

\[
\text{LoG}(X, \sigma) = \sum_i \frac{\partial^2 G(X)}{\partial x_i^2} = \frac{n}{\sigma^2} G(X) + \frac{1}{\sigma^2} ||X||^2 G(X), \tag{12}
\]

where \( G(X) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp(-\frac{||X||^2}{2\sigma^2}) \) is an \( n \)-dimensional Gaussian function. So, at each sample points, from Eqs. (11) and (12), we get the Laplacian as

\[
\nabla^2 V(y_i) = \sum_i u_i \text{LoG}(y_j - y_i, \sigma) = \frac{n}{\sigma^2} (\bar{u}_i - u_i), \tag{13}
\]

where \( \bar{u}_i = \sum_{i,j} ||y_j - y_i||^2 \) can be considered as a local average in some way. Other weighted average such as Gaussian can also be considered too.

### 3.1. Iterative solution

Using Eq. (13), we can rewrite the Euler–Lagrange equation (10) as

\[
E_x E_x u_i + E_y E_t - x^2 \frac{n}{\sigma^2} (\bar{u}_i - u_i) = 0. \tag{14}
\]

Solving \( u_i \) from Eq. (14), we can get the iterative solution:

\[
u_i^{n+1} = \bar{u}_i - (E_x E_x + x^2 I)^{-1} E_x (E_x u_i + E_t), \tag{15}
\]

where the constant coefficient \( \frac{n}{\sigma^2} \) is discarded without loss of generality.
To compute partial derivatives $E_X$ and $E_t$ robustly, we use local Gaussian smoothing to obtain the partial derivative of $E(X,t)$.

\[
E_X = \sum_i E(y_i)G_X(X-y_i,\sigma)
\]

\[
E_X = -\frac{1}{\sigma^2} \sum_i E(y_i)(X-y_i)^\top G(X-y_i,\sigma),
\]

(16)

\[
E_t = \sum_i \frac{E(y_i,t+\Delta t) + E(y_i,t)}{\Delta t} G(X-y_i,\sigma)
\]

(17)

where $G_X(X,\sigma) = -\frac{1}{\sigma} X^\top G(X,\sigma)$ is derivative of Gaussian, and $\bar{E}(y_i)$ is the average of the consecutive frames $\bar{E}(y_i) = \frac{\bar{E}(y_i,t+\Delta t) + \bar{E}(y_i,t)}{2}$.

3.2. Local smoothness constraint

Optical flow of grey-level image is a typical ill-posed problem. Smoothness constraint is often used to regularize this ill-posed problem. The smoothness constraint in multi-channel situation is more complicated than single channel. Here, we must determine whether it is ill-posed firstly, which depends on $E_X, m,$ and $n$, when $m < n$, that is, the number of channel less than the dimension, the problem is ill-posed and regularization is needed, but when $m \geq n$, the intensity constraint equation (6) trends to be over-determined and it is necessary to determine whether regularization is needed or not locally.

When considering 2D RGB images, $m > n$, some multi-channel optical flow methods discard smoothness constraint [12,2], and apply least square error method to obtain the solution $V = (E_X^\top E_X)^{-1} E_X^\top E_t$. This is in fact equivalent to letting $\varepsilon = 0$ in Eq. (15). However there are probably many regions where color channels are homogeneous and $E_X^\top E_X$ is nearly singular and regularization is needed to obtain a robust solution. Here, we smaller smoothness constraint is applied to reduce blurring effect at pixels with full rank matrix $E_X^\top E_X$ and larger smoothness constraint to obtain robust flow where $E_X^\top E_X$ is singular.

\[
\begin{cases} 
\varepsilon_{\min}^2 & \text{if } \text{rank}(E_X^\top E_X) = n, \\
\varepsilon_{\max}^2 & \text{if } \text{rank}(E_X^\top E_X) < n.
\end{cases}
\]

The singularity is determined by calculating the condition number. If condition number is above a predefined precision $\varepsilon$, the matrix is thought to be singular. In the following section $\varepsilon_{\min}^2$ is selected to be $\frac{1}{2} \varepsilon_{\max}^2$ in color image optical flow. When $m < n$, $E_X^\top E_X$ is always singular, and smoothness constraint is selected globally $\varepsilon = \varepsilon_{\max}^2$.

4. Scale-space pullback of color images

The above method can be applied to color images directly. Let $E_c = (e_1, e_2, e_3)^\top$ be the RGB representation of a color image. The intensity equation is still $E_c u + E_c v + E_c = 0$. Its iterative solution is given by:

\[
u^{n+1} = \tilde{\nu}^{n} - \frac{E_X^\top E_X [E_c \tilde{\nu}^{n} + E_c \tilde{\nu}^{n} + E_c]}{\varepsilon^2 + E_X^\top E_X + E_c^\top E_c},
\]

(19)

\[
u^{n+1} = \tilde{\nu}^{n} - \frac{E_X^\top E_X [E_c \tilde{\nu}^{n} + E_c \tilde{\nu}^{n} + E_c]}{\varepsilon^2 + E_X^\top E_X + E_c^\top E_c}.
\]

(20)

It resembles the single channel optical flow [6] only by substituting some scalar products with inner products and global $z$ with local $z$.

In scale-space optical flow, as the scale increases, image points in the original image trends to aggregate into several clusters. At a larger scale, each cluster center represents many points in original image while points between cluster centers have few corresponds in the original image, so the point near the cluster centers are more important. At each scale, we can build a flow density map $\mathcal{F}(X,t)$ by counting how many points in

Fig. 2. Flow and density: (a) original image, (b) accumulate translation at each pixel, (c) resultant image, (d) density map, in which the dominant white points are cluster centers.
the original image correspond to position \( X \) at current scale, i.e., each pixel in the original image votes for its corresponding pixel in the current image (see Fig. (2)). We use the density to emphasize the intensity constraint of dense points in image filtering which can be regarded as another type of \( \alpha \) trick. Using \( F(X,t) \) to weight the intensity constraint item in energy functional (9), we obtain the weighted energy functional

\[
J[V] = \int F(X,t)(E_X V + E_t)^T(E_X V + E_t) + \alpha^2 \\
\times \sum_{i,j} \left( \frac{\partial u_i}{\partial X_j} \right)^2 dX. \tag{21}
\]

The iterative solution becomes

\[
u_{l+1}^i = \bar{u}_l^i - F(X,t) \left[ F(X,t)E_X^T E_X + \alpha^2 I \right]^{-1} E_X^T (E_X \bar{u}_l^i + E_t).	ag{22}\]

In implementation, we introduced a new smoothness constraint as \( b^2(X,t) = \frac{\alpha^2}{\epsilon^2 + \alpha^2} \), where \( \epsilon \) is a tiny positive number to prevent divide-by-zero errors.

The color image scale-space optical flow is implemented based on general optical flow. First, we convert the image to Lu*v* color space [10] for a better color distance metric, then incrementally blur the Lu*v* image with Gaussian and compute the optical flow between current image and

---

**Fig. 3.** SSOF pullback results; left, original; middle, filtered image with global smoothness constraint; right, filtered image with local smoothness constraint.
blurred image. Several experimental results are demonstrated in Fig. 3, one can see that localized smoothness constraint can sharpen the boundary considerably, for example, the boundary of the wings of the butterflies.

5. Scale-space pullback of 3D images

We also applied our method on 3D images. Experiments are conducted on CT image of human knee and MR image of brain from the Stanford volume data archive [http://graphics.stanford.edu/data/voldata/]. The sample slices of the data and results are shown in (Fig. 4). From the results, one can see that the noise and the small structure are remove successfully without blurring large structure.

6. Conclusion

In this paper, we generalized Yang and Ma’s work to generic images. First, a generalized Horn–Schunck algorithm for computing optical flow under localized smoothness constraints was developed. Then, we introduced another local smoothness trick based on flow density. At last, we applied the proposed method to color image scale-space pullback. The experimental results on color images and 3D images demonstrated the validity of our methods. By removing small structures in a scale incremental manner, the proposed technique successfully avoids the awkward situation that when a scale parameter is small, the results at this scale will be useless if we decide to consider
another larger scale. The perceptual effect of any multiscale representation can be improved by this SSOF pull-back technique.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cviu.2006.07.005.

References