Optic flow based on multi-scale anchor point movement and discontinuity-preserving regularization


Eindhoven University of Technology, Den Dolech 2, Postbus 513, 5600 MB Eindhoven, The Netherlands

A R T I C L E   I N F O

Keywords:
Image processing
Optic flow
Multi-scale
Scale-space
Toppoints
Discontinuity-preserving regularization

A B S T R A C T

We introduce a new method to determine the flow field of an image sequence using multi-scale anchor points. These anchor points manifest themselves in the scale-space representation of an image. The novelty of our method lies largely in the fact that the relation between the scale-space anchor points and the flow field is formulated in terms of soft constraints in a variational method. This leads to an algorithm for the computation of the flow field that differs fundamentally from previously proposed ones based on hard constraints. We show a significant performance increase when our method is applied to the Yosemite image sequence, a standard and well-established benchmark sequence in optic flow research. Also, it is shown that this performance is not sensitive to slight changes in the two parameters used and that, with the same parameter values, our method yields very good results in the Rubber Whale image sequence as well.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Optic flow describes the apparent motion in an image sequence. A variety of approaches exists to estimate this motion. Survey papers include those by Barron et al. [1] and Mitchie et al. [2]. Differential methods are based on the most widespread approach, which uses spatiotemporal derivatives to describe the local image structure. The flow field is assumed to connect points in subsequent frames of the image sequence with similar structure. For example, in one of the earliest methods, proposed by Horn and Schunck [3], this “structure” is the image intensity, which leads to the well-known optic flow constraint equation (henceforth abbreviated OFCE). An overview of current developments in differential methods can be found in Bruhn et al. [4]. A problem that is encountered by these methods is that the structure does not always remain constant over time. For example, the global image intensity may vary over time. More complex terms to describe the structure can be used to overcome this problem [5,6]. A second problem is that many possible solutions exist, since points on level-sets have the same image intensity. This requires a so-called prior, which determines a unique solution based on prior knowledge. A prior usually is a regularization term, which can for example prefer an overall smooth solution with sparse discontinuities [7–9].

Another well performing approach is that of region matching, in which the image is split up into small blocks, each of which is translated to match the image neighborhood [10]. Because of their low computational cost, these methods are widely used in applications such as temporal up-scaling of video signals and video compression.

Our method can be placed in the category of feature-tracking methods. An overview of such methods can be found in [11]. However, in contrast to most feature-tracking algorithms, the features we use do not correspond to specific points in the image sequence. Instead, we use anchor points that exist at different scales in scale-space, called toppoints (properly defined in Section 2.1). Therefore, instead of corresponding to points, the features we track actually represent entire regions in the image sequence. Using toppoints to extract the motion from an image sequence has been first proposed by Janssen et al. [12] and Florack et al. [13]. In these papers, the relation between the toppoint velocity and the flow field was implemented using a hard constraint, which means that this constraint has to be fulfilled exactly. The advantage is that their method is entirely parameter free, but the price of this is sensitivity to outliers. In the method presented in this paper, a 1-parameter soft constraint is used, yielding higher robustness against errors in the estimated toppoint velocity or deviations from the stipulated relation between toppoint velocity and the flow field.

Toppoints are found throughout the scale-space of each frame of the image sequence as isolated entities. Therefore they are truly multi-scale, in contrast to other multi-scale features which are found by applying scale-selection to points that exist at every scale, such as corners. Another use of toppoints is to reconstruct
an image from the values of derivatives taken at toppoint positions, cf. the papers by Liljholm [14], Nielsen [15] and Jansen [16]. In these papers it is shown that features at the toppoint positions can be used to efficiently represent the information contained in an image. An important property is that the amount of toppoints found in a certain area of the image is proportional to the amount of information in that area.

2. Theory

In this chapter we will first explain how the scale-space representation of an image is defined and what toppoints are. Also important properties of toppoints are mentioned and we try to give toppoints a more intuitive meaning with some visualizations. Next we explain how to calculate toppoint velocities, and the method used to obtain the actual flow field from the toppoint velocities.

2.1. Scale-space and toppoints

The scale-space representation \( f_0(x,y,s) \), where \( f \in C^\infty(\mathbb{R}^2 \times \mathbb{R}^+) \), of a static scalar image \( f_0 \in L_2(\mathbb{R}^2) \) is defined by the convolution of the image with a Gaussian kernel \( \phi_s(x,y) = \phi(x,y,s) \), where \( \phi \in C^\infty(\mathbb{R}^2 \times \mathbb{R}^+) \) and \( s \in \mathbb{R}^+ \) denotes the scale on scale-space (for tutorial books on scale-space see ter Haar Romeny [17], Florack [18] and Lindeberg [19]):

\[
f : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R} : (x,y,s) \mapsto f(x,y,s) \equiv f_0 * \phi_s(x,y),
\]

\[\phi_s(x,y) = \phi(x,y,s) = \frac{1}{4\pi s^2} \exp\left(-\frac{x^2+y^2}{4s^2}\right).\]  \hfill (1)

This results in a 3D function, where a slice of constant scale represents a blurred version of the original image.

The scale-space of an image fulfills the heat equation, since after the convolution product of a derivative of the image \( f_0 \) with a Gaussian filter \( \phi_s \), see Eq. (1), using the property that \( \phi_s \) is a Schwartz function:

\[
\frac{\partial f_0}{\partial t} = \nabla^2 f_0 \equiv \phi_s * \frac{\partial^2 f_0}{\partial x^2} + \phi_s * \frac{\partial^2 f_0}{\partial y^2} = \phi_s * \frac{\partial^2 f_0}{\partial x^2} + \phi_s * \frac{\partial^2 f_0}{\partial y^2} = \phi_s * \frac{\partial^2 f_0}{\partial x^2} + \phi_s * \frac{\partial^2 f_0}{\partial y^2}.
\]  \hfill (4)

In fact, because \( f_0 \) is often not \((m+n)\) times differentiable, we define the scale-space of an image derivative by the right hand side of Eq. (4). This results in a lower-bound on the scale at which derivatives can be calculated numerically, which increases with derivative order. Derivatives with respect to scale can be calculated using only spatial derivatives by means of Eq. (2).

2.2. Toppoint velocity

If we consider a sequence of successive images, or a movie, in which objects move, the toppoints will move as well. The movement of toppoints in spatial and scale direction is defined as: \((x,y,s) \in \mathbb{R}^3\). Note that e.g. \( \dot{x}(t) = \partial_x x(t) \) represents the time derivative of the \( x(t) \) position of the toppoint. An expression for this toppoint movement can be obtained by implicitly differentiating the definition of toppoints as stated in Eq. (3) with respect to the time parameter \( t \):

\[
d \frac{d}{dt} \left[ \begin{array}{c} \frac{\partial x(t)}{\partial t} \\ \frac{\partial y(t)}{\partial t} \\ \frac{\partial s(t)}{\partial t} \end{array} \right] = \left[ \begin{array}{ccc} 0 & 0 & 0 \\ \frac{\partial^2 x(t)}{\partial x^2} & \frac{\partial^2 y(t)}{\partial x^2} & \frac{\partial^2 s(t)}{\partial x^2} \\ \frac{\partial^2 x(t)}{\partial y^2} & \frac{\partial^2 y(t)}{\partial y^2} & \frac{\partial^2 s(t)}{\partial y^2} \end{array} \right].
\]  \hfill (5)

If the matrix is invertible, Eq. (5) supplies us with a scheme to calculate the movement of toppoints in an image sequence. The
notation for derivatives of $\text{det } H$ is abbreviated to avoid cumbersome notation. When we expand $\partial \text{det } H$ for example, we obtain (using Eq. (2) to express scale derivatives in spatial derivatives):

$$
\partial_x (f_{x}^2 f_{y}^2 - f_{x}^2 f_{y}^2) = f_x(f_{x}^2 + f_{y}^2) + f_y(f_{x}^2 + f_{y}^2) - 2 f_{x} f_{y} f_{x} f_{y}.
$$

An estimation of the position of toppoints can be used to find a more accurate location. Florack and Kuijper [22,23] developed a method that iteratively refines the estimated position, $x_0 = \{x_0, y_0, z_0\} \in \mathbb{R}^3 \times \mathbb{R}^+$. To the true toppoint position, $x_0 = \{x_0, y_0, z_0\} \in \mathbb{R}^3 \times \mathbb{R}^+$. To acquire a dense 2D flow field movement of the toppoints from one frame to the next.

We can write this system of equations in matrix form:

$$
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
= 
\begin{bmatrix}
f_x & f_y & f_z \\
f_x & f_y & f_z \\
\partial_x \text{det } H & \partial_y \text{det } H & \partial_z \text{det } H
\end{bmatrix}
\begin{bmatrix}
x_0 - x_0 \\
y_0 - y_0 \\
\gamma_0 - \gamma_0
\end{bmatrix},
\tag{9}
$$

where image derivatives are calculated at the estimated toppoint position $x_0$. This gives us the following scheme to refine toppoint positions:

$$
\begin{bmatrix}
x_0 = x_0 - \\
y_0 = y_0 - \\
\gamma_0 = \gamma_0 - 
\end{bmatrix}
\begin{bmatrix}
f_x \\
f_y \\
f_z
\end{bmatrix}
\begin{bmatrix}
\partial_x \text{det } H & \partial_y \text{det } H & \partial_z \text{det } H
\end{bmatrix}^{-1}
\begin{bmatrix}
f_x \\
f_y \\
f_z
\end{bmatrix}.
\tag{10}
$$

Since we approximate the local image structure with a first order Taylor polynomial, applying Eq. (10) will not give the exact position of the toppoint. However, the calculation can be executed iteratively for increasing precision. Again, although notation suggests otherwise, no scale-derivatives need to be taken explicitly by virtue of the heat equation (Eq. (2)).

Using the estimated toppoint velocity, we estimate the toppoint position in the next frame of the image sequence. Consequently, the position of the toppoint in the next frame is refined with the iterative method explained above. This refined position is used to calculate a more accurate estimation of the movement of the toppoints from one frame to the next.

2.3 Optic flow using toppoints

The velocity of the scale-space toppoints forms a sparse 3D flow field. In optic flow, the goal is to acquire a dense 2D flow field which describes the velocity in each pixel of the image sequence. In order to obtain the dense 2D flow field from the sparse 3D one the following assumption is made:

2.3.1 Assumption 1

The velocity of toppoints in the scale-space of the image corresponds to the values at those points in the scale-spaces of $u(x,y)$ and $v(x,y)$:

$$
\langle u, \phi_i \rangle = U_i, \\
\langle v, \phi_i \rangle = V_i,
\tag{11}
$$

where $U_i \in \mathbb{R}$ and $V_i \in \mathbb{R}$ are obtained by applying Eq. (5) at the toppoint positions of the image sequence, and $\phi_i$ are Gaussian functions shifted to spatial position $x_i,y_i$ and with scale $s_i$ (recall Eq. (1)). Here and henceforth, $\langle \ldots \rangle$ indicates a standard $L_2$ inner product.

This assumption alone does not uniquely determine the flow field. Therefore, we use a flow driven isotropic prior, which allows for some discontinuities in the flow field. We combine this prior with the assumption regarding toppoint velocities in the following energy functional:

$$
E(u,v) = \int \frac{\gamma}{2} \sqrt{|\nabla u|^2 + |\nabla v|^2} + c^2 + \sum_{i=1}^{N} \left( \langle u, \phi_i \rangle - U_i \right)^2 + \left( \langle v, \phi_i \rangle - V_i \right)^2 d\Omega,
\tag{12}
$$

where $c$ is a contrast parameter, $\gamma$ determines the smoothness of the resulting flow field and $N$ denotes the number of toppoints. Minimizing this energy functional leads to the dense flow field. Using variational calculus, we obtain the Euler–Lagrange equations corresponding to this energy functional. These are discretized using $\beta$-splines [24] and the resulting system of equations is solved using the BiCG-Stab algorithm [25].

Besides the toppoints of the regular image, we also calculate the movement of higher-order structures in the image, such as edges.

3. Numerical evaluation

The error measure for flow fields used in literature is the angular error, as first proposed by Fleet and Jepson [26]. This measure describes the angle between the estimated 3D flow vector $\mathbf{v}_e = [u_e, v_e, 1]$ and the true flow vector $\mathbf{v}_t = [u_t, v_t, 1]$. In order to objectively compare different methods, the average angular error, or AAE, is used.

Fig. 2 shows the Yosemite image sequence, which is used in optic flow literature as a benchmark sequence. This sequence tests multiple aspects of the performance of optic flow methods: it contains spatial discontinuities, brightness change (the sky increases in brightness), rigid and non-rigid transformations.

The second image sequence for which we show the results is the Rubber Whale image sequence, of which one frame together with the ground truth flow field is shown in Fig. 3.

In Table 1 different methods that introduced significant novelties can be found together with the improvement of the AAE of the Yosemite image sequence since Horn and Schunck introduced their method in 1981.

The flow field that is obtained by our method for the Yosemite image sequence can be found in Fig. 6, together with the angular error and toppoint locations. We can see that, apart from the discontinuity at the border between the landscape and the sky, the flow field is fairly accurate, albeit not state-of-the-art. This discontinuity between the landscape and the sky results in the largest error. This is partially caused by the low number of toppoints found in the low-texture sky, and partially by the suboptimal choice of prior. Using a smoothness term with better discontinuity-preserving properties may improve this result significantly.

Another observation is that the motion of the toppoints in the landscape seems to extend into the sky, since the toppoint density in the landscape is much higher and all toppoints are given equal weight in the energy functional that is minimized. Because the flow field at the sky is completely uniform, it could also be modeled very well as motion caused by camera pan.

The flow field of the Rubber Whale image sequence, shown in Fig. 6, shows similar characteristics. The error in most areas is
very low, but there seem to be less (stable) toppoints around the discontinuities in the flow field, resulting in the largest error in these areas.

The AAE we obtained for different values of parameters $\varepsilon$ and $\gamma$ for both image sequences is shown in Table 2 and Figs. 4 and 5. As can be seen, a wide range of parameter values yields comparable results. This is an indication of the stability of the method, which is important since other image sequences might require different parameter values for optimal performance.

The optimal performance of the method results in an AAE of 4.80° for the Yosemite sequence and 8.50° for the Rubber Whale sequence. For an $\varepsilon$ value of 9 and a $\gamma$ value of $2 \times 10^6$ the method shows a performance very close to the optimum for both image sequences, with an AAE of 4.81° for the Yosemite sequence and 8.53° for the Rubber Whale sequence. Also, as can be seen in these figures, there seems to be a hyperbolic relation between $\varepsilon$ and $\gamma$ for the optimal solution (Fig. 7).

4. Conclusion and future work

We have shown that the information toppoint movement provides a fairly accurate estimation of the flow field of an image sequence. We obtained a flow field with an AAE as low as 4.80° for the Yosemite sequence, even without the use of a complex, parameter-rich and computationally expensive method to preserve discontinuities, such as level-sets. In comparison: the method proposed in [12], using hard constraints, resulted in an AAE of 19.19°.

This preliminary result is promising for several reasons: (i) Unlike superior sophisticated methods our method is characterized by only two global parameters ($\varepsilon$ and $\gamma$). (ii) Toppoint representations are typically very sparse (1837 toppoints for the 79,316 pixels in the Yosemite sequence). (iii) The method itself can be easily modified so as to account for different or additional anchor points and more effective priors.

The principle novelty of our approach is the term in the energy functional which provides the information on the flow field. Many improvements in differential methods have been made in the regularization term, which can also be incorporated into our

---

**Table 1**

Yosemite sequence results of other methods.

<table>
<thead>
<tr>
<th>Technique</th>
<th>AAE (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original, only smoothness [3]</td>
<td>32.43</td>
</tr>
<tr>
<td>Region matching [27]</td>
<td>15.84</td>
</tr>
<tr>
<td>Coarse-to-fine approach [28]</td>
<td>13.16</td>
</tr>
<tr>
<td>Discontinuity preservation [7]</td>
<td>11.71</td>
</tr>
<tr>
<td>Improvement of Nagel [8]</td>
<td>5.53</td>
</tr>
<tr>
<td>Spatio-temporal smoothness [29]</td>
<td>4.85</td>
</tr>
<tr>
<td>Monogenic curvature tensor constancy [6]</td>
<td>2.67</td>
</tr>
<tr>
<td>Piecewise smoothness with level-sets [30]</td>
<td>1.78</td>
</tr>
<tr>
<td>High order data term, 3D smoothness [5]</td>
<td>1.64</td>
</tr>
<tr>
<td>Same as Papenberg, with level-sets [31]</td>
<td>0.92</td>
</tr>
</tbody>
</table>

The first four results are obtained from Barron et al. [1].

---

Please cite this article as: P.A.G. van Dorst, et al., Optic flow based on multi-scale anchor point movement and discontinuity-preserving regularization, Pattern Recognition (2010), doi:10.1016/j.patcog.2010.09.003
method. Also other anchor points can be added, such as SIFT feature points [32].

Other possible improvements are the addition of weights to the toppoint constraints. These could for example be dependent on a stability measure for the toppoint position, or an accuracy measure for the toppoint movement. Also we have seen that information on the motion of toppoints is extended from areas of high toppoint density into areas of low toppoint density, since the toppoint weight is equal for all toppoints. When some sort of measure for toppoint density is developed, this could be used to compensate for this bias.

Since the Yosemite image sequence is the benchmark sequence used in literature, for most methods only the AAE of this sequence is available. Our approach is very robust compared to differential methods, since it is inherently multi-scale and less sensitive to changing brightness, and rather generic, as it requires

<table>
<thead>
<tr>
<th>γ</th>
<th>1 × 10^6</th>
<th>2 × 10^6</th>
<th>3 × 10^6</th>
<th>4 × 10^6</th>
<th>5 × 10^6</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε = 1</td>
<td>7.09/11.14</td>
<td>6.23/10.46</td>
<td>5.75/10.46</td>
<td>5.48/10.03</td>
<td>5.44/9.73</td>
</tr>
<tr>
<td>ε = 2</td>
<td>6.16/10.65</td>
<td>5.47/10.03</td>
<td>5.15/9.51</td>
<td>4.98/9.21</td>
<td>5.35/9.01</td>
</tr>
<tr>
<td>ε = 3</td>
<td>5.75/10.46</td>
<td>5.15/9.52</td>
<td>4.95/9.11</td>
<td>4.84/8.88</td>
<td>4.81/8.73</td>
</tr>
<tr>
<td>ε = 4</td>
<td>5.46/10.05</td>
<td>4.98/9.21</td>
<td>4.84/8.70</td>
<td>4.80/8.70</td>
<td>4.83/8.60</td>
</tr>
<tr>
<td>ε = 5</td>
<td>5.28/9.75</td>
<td>4.89/8.73</td>
<td>4.81/8.73</td>
<td>4.83/8.60</td>
<td>4.89/8.54</td>
</tr>
<tr>
<td>ε = 7</td>
<td>5.05/9.35</td>
<td>4.81/8.78</td>
<td>4.84/8.58</td>
<td>4.94/8.52</td>
<td>5.08/8.50</td>
</tr>
<tr>
<td>ε = 8</td>
<td>4.98/9.22</td>
<td>4.80/8.70</td>
<td>4.88/8.52</td>
<td>5.02/8.50</td>
<td>5.19/8.57</td>
</tr>
</tbody>
</table>

Fig. 4. The AAE our method achieved for the Yosemite sequence with a range of values for the two parameters ε and γ.

Fig. 5. The AAE our method achieved for the Rubber Whale sequence with a range of values for the two parameters ε and γ.

Fig. 6. (Top) The flow field of the Yosemite image sequence calculated using our method. (Bottom) The angular error of the flow field calculated using our method, displayed as shades of grey, with white = 0° and black = 102.7°. Red dots indicate toppoints, the size of which is proportional to the scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: P.A.G. van Dorst, et al., Optic flow based on multi-scale anchor point movement and discontinuity-preserving regularization, Pattern Recognition (2010), doi:10.1016/j.patcog.2010.09.003
Fig. 7. (Top) The flow field of the Rubber Whale image sequence calculated using our method. (Bottom) The angular error of the flow field calculated using our method, displayed as shades of grey, with white = 0° and black = 121.9°. Red dots indicate topknots, the size of which is proportional to the scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

only two global regularity parameters. Therefore it is expected to perform well on more challenging image sequences, such as those with opacity, reflections or a significant amount of noise. This is the subject of further research.

However, the multi-scale nature seems to perform less accurately around discontinuities in the flow field, such as that around the border between the sky and terrain in the Yosemite sequence and the different objects in the Rubber Whale sequence.

As we have seen indications of a hyperbolic relation between the optimal values of $\varepsilon$ and $\gamma$, it could be fruitful to devote future research to the exact influence of these parameters and how to estimate a proper value for a wide range of different image sequences.

References