Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Robust optical flow estimation based on brightness correction fields*

Wei WANG†[1], Zhi-xun SU†[1], Jin-shan PAN[1], Ye WANG[2], Ri-ming SUN[1]

([1]*School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China*)

([2]*Department of Mathematics, Harbin Institute of Technology, Harbin 150006, China*)

†E-mail: garywangzi@gmail.com; zxsu@dlut.edu.cn

Received Mar. 17, 2011; Revision accepted June 17, 2011; Crosschecked Nov. 4, 2011

Abstract: Optical flow estimation is still an important task in computer vision with many interesting applications. However, the results obtained by most of the optical flow techniques are affected by motion discontinuities or illumination changes. In this paper, we introduce a brightness correction field combined with a gradient constancy constraint to reduce the sensibility to brightness changes between images to be estimated. The advantage of this brightness correction field is its simplicity in terms of computational complexity and implementation. By analyzing the deficiencies of the traditional total variation regularization term in weakly textured areas, we also adopt a structure-adaptive regularization based on the robust Huber norm to preserve motion discontinuities. Finally, the proposed energy functional is minimized by solving its corresponding Euler-Lagrange equation in a more effective multi-resolution scheme, which integrates the twice downsampling strategy with a support-weight median filter. Numerous experiments show that our method is more effective and produces more accurate results for optical flow estimation.

Key words: Optical flow field, Variational methods, Brightness correction fields, Median filter, Multi-resolution
doi:10.1631/jzus.C1100062 **Document code:** A **CLC number:** TP391.4

1 Introduction

Estimating a displacement field for consecutive images from an image sequence is one of the major challenges in computer vision. It arises whenever one aims to identify correspondences between points in pairs of images. Examples include motion estimation (Gilland *et al.*, 2008), tracking (Dessauer and Dua, 2010), and medical multi-modal registration (Wang *et al.*, 2007). The resulting dense correspondence between pairs of points in either image can subsequently be used for the structure-from-motion algorithm (Fakih and Zelek, 2008), object recognition (Efros *et al.*, 2003), and other higher level tasks.

Starting with the seminal work of Horn and Schunck (1981), most of state-of-the-art methods are based on minimizing the energy

$$E(\boldsymbol{U}) = E_{\mathrm{d}}(\boldsymbol{I}_1, \boldsymbol{I}_2, \boldsymbol{U}) + \lambda E_{\mathrm{s}}(\boldsymbol{U}), \qquad (1)$$

where the unknown optical flow field $\boldsymbol{U} : \Omega \to \mathbb{R}^2$, and two input images $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ are both defined on the image domain $\Omega \subset \mathbb{R}^2$. The data term $E_{\mathrm{d}}(\boldsymbol{I}_1, \boldsymbol{I}_2, \boldsymbol{U})$ measures the similarity of the input images for a given optical flow field, and the regularization term $E_{\mathrm{s}}(\boldsymbol{U})$ allows to impose some prior on the optical flow field $\boldsymbol{U}$. Although a lot of work has focused on improving energy models and optimizing algorithms, there is still room for improvement.

For the data term, the original energy model of optical flow relies on the brightness constancy constraint. Any changes of the illumination in the scene violating the brightness constancy assumption lead

* Project supported by the National Natural Science Foundation of China (No. U0935004) and an IDeA Network of Biomedical Research Excellence (INBRE) grant from the National Institutes of Health (NIH) (No. 5P20RR01647206)

to unsatisfactory results. To deal with the problem of brightness variations, many methods have been proposed (Gennert and Negahdaripour, 1987; Negahdaripour, 1998; Haussecker and Fleet, 2001; Kim *et al.*, 2005; Teng *et al.*, 2005). For example, Gennert and Negahdaripour (1987) proposed to replace this constancy constraint with a linear model between image brightness values. Haussecker and Fleet (2001) proposed some physical models of brightness variations, including the brightness variations caused by diffusion, moving illumination envelope or changed surface orientation, etc., for optical flow computation. However, these physical models are usually constructed from some prior knowledge about the system. Thus, their methods are appropriate only for the applications where the physical model is known in advance. Also, they cannot deal with the optical flow globally. Kim *et al.* (2005) proposed an integrated approach that incorporates Gennert and Negahdaripour (1987)'s formulation with the robust statistical framework of Black and Anandan (1996). Teng *et al.* (2005) adopted the revised definition of the optical flow proposed by Gennert *et al.* and embedded a generalized dynamic image model (GDIM) (Negahdaripour, 1998) into a gradient-based regularization method to estimate the revised optical flow. Although these methods can deal with the illumination variations, they have many parameters that need to be carefully tuned in order to obtain desirable results. They adopt the robust log function, leading to a highly non-convex model, and this adds to the difficulty in the computing.

For the regularization term, the classic Horn and Schunck (HS) model uses a homogeneous regularization term that does not respect any flow discontinuities. Since different image objects may move in different directions, it is also desirable to permit discontinuities. The total variation (TV) regularization allows to retrieve flow discontinuities. There are also many works on robust functions and isotropic/anisotropic regularization formulations to retrieve flow discontinuity (Shulman and Herve, 1989; Black and Anandan, 1996; Brox *et al.*, 2004; Bruhn *et al.*, 2005; Zach *et al.*, 2007; Steinbrücker and Pock, 2009). The TV regularization is an L^1 penalization of the flow gradient magnitudes. However, due to the tendency of the L^1 norm to favor sparse solutions (i.e., lots of 'zeros'), the fill-in effect leads to piecewise constant solutions ('staircasing' in a 1D setting) in weakly textured areas. Huber (1973) proposed a quadratic penalization function in the field of robust statistics to reduce 'staircasing' effect. Werlberger *et al.* (2009) proposed to replace an isotropic TV regularization, e.g., used in Wedel *et al.* (2009a), with an anisotropic (image-driven) Huber regularization term. For many real-world examples, discontinuities of the motion field tend to coincide with object boundaries and discontinuities of the brightness function. Wedel *et al.* (2009b) proposed a structure-adaptive regularization term.

The optimization algorithm and the implementation are also important for improving the accuracy of the estimated flow fields. Most methods adopt the coarse-to-fine strategy. These methods include: temporal averaging of image derivatives (Hsiao *et al.*, 2003), texture decomposition (Zach *et al.*, 2007; Wedel *et al.*, 2009a), the robust statistical framework of Black and Anandan (1996), and bicubic interpolation-based warping (Lempitsky *et al.*, 2008). One problem of the coarse-to-fine strategy resides in the warping process, which transforms images by a coarser estimated field. Works of Cassisa *et al.* (2009) and Steinbrücker and Pock (2009) for optical flow computation do not need the warping step. If the estimation at a coarser level is pretty close to the exact coarser field, however, errors introduced by the warping process are negligible. Thus, both the downsampling strategy and using an advanced filter to intermediate flow values before the warping process are very important. Wedel *et al.* (2009a) adopted the median filtering after each incremental estimation step to remove a lot of outliers. Sun *et al.* (2010) used a weighted filter (Yoon and Kweon, 2006) model to improve the accuracy of the recovered optical flow field.

Based on the above analysis, we propose a new energy model for robust optical flow estimation. Its data term combines the brightness correction field with gradient constancy constraints. Unlike the physical model and the linear model of brightness variations mentioned above, this brightness correction field is based on a regularization operator. It takes into account non-stationary brightness relationship and reduces the sensibility to brightness changes. To decrease the 'staircasing' effect caused by the TV regularization term in weakly textured areas, a structure-adaptive regularization term based on the Huber norm is adopted. The Euler-Lagrange

equation of the proposed energy model is solved by an effective multi-resolution scheme, which uses the downsampling strategy twice with different downsampling factors. Thus, the proposed strategy can recover the missing details that are caused by the traditional coarse-to-fine scheme. After each warping step, a support-weight median filter is applied to the newly computed flow field. It helps to remove outliers robustly.

2 Methodology

2.1 Data term with the brightness correction field

At the beginning of optical flow estimation, the data term is based on the assumption that the brightness of a pixel is not changed by the displacement. However, in realistic scenarios, the brightness constancy assumption almost never holds. For example, a pixel can change its brightness value because an object moves (translates or rotates) to another part of the scene with the different lighting or because the illumination of the scene changes (globally or locally) in time. To handle this issue, one has to take into account non-stationary brightness variation relationship. Previous works have focused on the physical model (Haussecker and Fleet, 2001) or the linear model of brightness variations (Gennert and Negahdaripour, 1987; Kim *et al.*, 2005; Teng *et al.*, 2005). Unlike these two models of the brightness variation, our brightness correction field model does not need to know in advance the physical model underlying the brightness variation and does not have many parameters that need to be carefully tuned in order to obtain desirable results. It is directly defined as

$$\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U}(\boldsymbol{X}))=\boldsymbol{I}_1(\boldsymbol{X})-\boldsymbol{M}(\boldsymbol{X}), \tag{2}$$

where $\boldsymbol{M}(\boldsymbol{X})$ is a brightness correction field (note that for any two images there always exists a correction field $\boldsymbol{M}(\boldsymbol{X})$). $\boldsymbol{U}(\boldsymbol{X})$ is the optical flow vector representing the displacement between frames $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$. The maximum a posteriori (MAP) approach to estimating $\boldsymbol{U}$ and $\boldsymbol{M}$ is to maximize the conditional probability:

$$p(\boldsymbol{U},\boldsymbol{M}|\boldsymbol{I}_1,\boldsymbol{I}_2)\propto p(\boldsymbol{I}_1,\boldsymbol{I}_2|\boldsymbol{U},\boldsymbol{M})p(\boldsymbol{U})p(\boldsymbol{M}). \tag{3}$$

Here the independence of $\boldsymbol{M}$ and $\boldsymbol{U}$ is assumed. The term $p(\boldsymbol{I}_1,\boldsymbol{I}_2|\boldsymbol{U},\boldsymbol{M})$ is a joint likelihood of the images, which leads to the familiar similarity measure of the brightness value constancy assumption. $p(\boldsymbol{U})$ is a prior used to regularize the optical flow $\boldsymbol{U}$ and $p(\boldsymbol{M})\propto\exp(-\beta||\boldsymbol{P}\boldsymbol{M}||^2)$ is a prior on $\boldsymbol{M}$ that reflects our assumption on spatial brightness interactions, where $\boldsymbol{P}$ is a regularization operator for $\boldsymbol{M}$ and β is the equilibrium parameter of two terms. It is also assumed that pixel-wise probabilities are independent and identically distributed, but only when given the correction field. MAP estimation in Eq. (3) is equivalent to the minimization of the following objective function:

$$\begin{aligned}E_{\mathrm{db}}(\boldsymbol{U},\boldsymbol{M})=&||\boldsymbol{I}_1(\boldsymbol{X})-\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})-\boldsymbol{M}||^2\\&+\beta||\boldsymbol{P}\boldsymbol{M}||^2,\end{aligned} \tag{4}$$

where images $\boldsymbol{I}_1$, $\boldsymbol{I}_2$ and the correction field $\boldsymbol{M}$ are in matrix form, and $\|\cdot\|$ is Euclidean norm. Eq. (4) can also include the regularization on $\boldsymbol{U}$, because it depends on the $\boldsymbol{U}$ that comes from a prior $p(\boldsymbol{U})$ (Myronenko and Song, 2009). To make our data term more robust, the gradient constancy assumption is considered. Thus, the data term of our energy model is defined as

$$E_{\mathrm{d}}(\boldsymbol{U},\boldsymbol{M})=E_{\mathrm{db}}(\boldsymbol{U},\boldsymbol{M})+\mu E_{\mathrm{dg}}(\boldsymbol{U}), \tag{5}$$

where $E_{\mathrm{dg}}(\boldsymbol{U})=\sum_{\boldsymbol{X}}||\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})-\nabla\boldsymbol{I}_1(\boldsymbol{X})||_1$, $||\cdot||_1$ is L^1 norm, and μ is a tuning weight parameter.

2.2 Structure-adaptive regularization term based on the Huber norm

Since different image objects may move in different directions, it is desirable to permit discontinuity. TV regularization takes this discontinuity into account; however, it leads to the piecewise constant of estimated flow fields in weakly textured area. The piecewise constant effect can be reduced significantly by using a quadratic penalty: for small gradient magnitudes the estimated flow field sticks to linear penalty, while for larger magnitudes the estimated flow field maintains the discontinuity preserving properties known from TV. Huber (1973)'s function is an effective quadratic penalization in the field of robust statistics. Shulman and Herve (1989) were the first to apply it in the context of flow estimation. Werlberger *et al.* (2009) proposed an anisotropic (image-driven) Huber regularization term to replace the isotropic TV regularization. Meanwhile, for

many real-world examples, discontinuities of the motion field tend to coincide with object boundaries and discontinuities of the brightness function. Thus, Wedel *et al.* (2009b) proposed a structure-adaptive regularization term. Inspired by these methods, our regularization term is defined as

$$E_{\mathrm{s}}(\boldsymbol{U})=\sum_{\boldsymbol{X}}||w(\boldsymbol{X})\nabla\boldsymbol{U}(\boldsymbol{X})||_{\delta}, \tag{6}$$

$$||j||_{\delta}=\begin{cases}|j|^2/(2\delta), & |j|\leq\delta,\\ |j|-\delta/2, & \text{otherwise},\end{cases} \tag{7}$$

where the threshold δ is set to 0.01, $w(\boldsymbol{X})=\exp(-||\nabla\boldsymbol{I}_1||^k)$, and k is set to 0.7 in our experiments. $w(\boldsymbol{X})$ is the simple structure adaptive map that maintains motion discontinuity (Wedel *et al.*, 2009b).

Thus, our objective function is defined as

$$E(\boldsymbol{U},\boldsymbol{M})=E_{\mathrm{d}}(\boldsymbol{U},\boldsymbol{M})+\lambda E_{\mathrm{s}}(\boldsymbol{U}), \tag{8}$$

where λ is a regularization weight parameter.

2.3 Formulation of the computing model

Eq. (8) is an energy function about the brightness correction field $\boldsymbol{M}$ and optical flow $\boldsymbol{U}$. $\boldsymbol{M}$ and $\boldsymbol{U}$ can be found by minimizing Eq. (8). We first solve analytically for $\boldsymbol{M}$ with fixed $\boldsymbol{U}$. Letting the derivative of Eq. (8) only with respect to $\boldsymbol{M}$ be zero, we obtain

$$\boldsymbol{M}=(\boldsymbol{I}+\beta\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})^{-1}\boldsymbol{r}, \tag{9}$$

where $\boldsymbol{r}=\boldsymbol{I}_1(\boldsymbol{X})-\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})$, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{r}$ is the residual image. $\boldsymbol{P}$ is a matrix, and is also the discrete derivative-based regularization of the brightness correction filed. The matrix $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$ is square, symmetric, and positive semi-definite. Thus, it allows spectral decomposition: $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{D}^{\mathrm{T}}, \boldsymbol{\Lambda}=\mathrm{diag}\{\lambda_1,\lambda_2,\cdots,\lambda_N\}, \lambda_i\geq 0, i=1,2,\cdots,N$, where N is the number of image pixels. Substituting $\boldsymbol{M}$ and $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$ back into Eq. (8), we obtain

$$E(\boldsymbol{U})=\boldsymbol{r}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{L}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{r}+\mu E_{\mathrm{dg}}(\boldsymbol{U})+\lambda E_{\mathrm{s}}(\boldsymbol{U}), \tag{10}$$

where $\boldsymbol{L}=\mathrm{diag}\{l_1,l_2,\cdots,l_N\}$, $l_i=\beta\lambda_i/(1+\beta\lambda_i)$, $0\leq\lambda_i\leq 1, i=1,2,\cdots,N$.

According to the analysis of Myronenko and Song (2009), $l_i=1/((\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{r})^2/\alpha+1)$, where $\boldsymbol{D}=[\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_N]$, and α is a trade-off parameter. The minimization problem of Eq. (10) is rewritten as

$$\begin{aligned}E(\boldsymbol{U})=&\sum_{n=1}^{N}\log\left[(\boldsymbol{d}_n^{\mathrm{T}}\boldsymbol{r})^2/\alpha+1\right]+\mu E_{\mathrm{dg}}(\boldsymbol{U})\\&+\lambda E_{\mathrm{s}}(\boldsymbol{U}).\end{aligned} \tag{11}$$

Note that the matrix $\boldsymbol{D}$ composed of basis eigenvectors is still unknown and thus needs to be defined. Here, the discrete cosine transform (DCT) basis (Strang, 1999) is used as a form of $\boldsymbol{D}$. The reasons are twofold: DCT eigenvectors correspond to the discrete derivative-based regularization $\boldsymbol{P}$; and, they are the eigenvectors of the covariance matrix of a weakly stationary stochastic process. It is well known in signal processing theory that with the assumption of the constant mean and the Toeplitz covariance, a stochastic process is said to be weakly stationary (Gray, 2006). Thus, using the DCT basis $\boldsymbol{D}$ is implicitly related to the assumptions of a weakly stationary residual image or a finite-order Markov process. In Eq. (10), the matrix multiplication $\boldsymbol{D}^{\mathrm{T}}\boldsymbol{r}$ is just a discrete cosine transformation of $\boldsymbol{r}$, which can be computed by DCT in $O(N\log N)$ (Strang, 1999).

$E(\boldsymbol{U})$ in Eq. (11) is considered as an energy functional with respect to the optical flow field $\boldsymbol{U}(\boldsymbol{X})$, and minimized by solving its corresponding Euler-Lagrange equation. Since $\varphi(x)=|x|$ is not continuously differentiable, we use $\varphi(x^2)=\sqrt{x^2+\varepsilon}, \varepsilon>0$ to approximate it. According to the calculus of variations (see the Appendix for the derivation of Eq. (12)), the Euler-Lagrange equation of Eq. (11) is obtained as follows:

$$\begin{aligned}-&\mathrm{idct}(\frac{2\boldsymbol{A}}{\boldsymbol{A}^2+\alpha})\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})\\&+\mu\nabla\varphi(||\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})-\nabla\boldsymbol{I}_1(\boldsymbol{X})||_1^2)\\&-\lambda\mathrm{div}(\phi'(\nabla\boldsymbol{U})\nabla\boldsymbol{U})=0,\end{aligned} \tag{12}$$

where $\phi(\nabla\boldsymbol{U})=||w(\boldsymbol{X})\nabla\boldsymbol{U}||_{\delta}$ and $\boldsymbol{A}=\mathrm{dct}(\boldsymbol{I}_1(\boldsymbol{X})-\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U}))$. The dct and idct are discrete cosine transform and inverse discrete cosine transform, respectively.

3 Implementation

Since Eq. (12) is highly non-linear, the minimization is not trivial. We use the first-order Taylor expansion to approximate the non-linear term $\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})$ and $\boldsymbol{A}$, i.e., $\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})=\nabla\boldsymbol{I}_2(\boldsymbol{X})+$

$\nabla^2 \boldsymbol{I}_2(\boldsymbol{X})^{\mathrm{T}}\boldsymbol{U}$ and $\boldsymbol{A}(\boldsymbol{X}+\boldsymbol{U}) = \boldsymbol{A}(\boldsymbol{X})+\nabla\boldsymbol{A}(\boldsymbol{X})^{\mathrm{T}}\boldsymbol{U}$. The linear system of equations can be solved by fixed point iterations.

To find the global optimum of the proposed energy, the multi-resolution method is commonly used. One problem of the coarse-to-fine strategy resides in the warping process, which transforms images by a coarser estimated field. However, if the estimation at the coarser level is pretty close to the exact coarser field, errors introduced by the warping process are negligible. To resolve this issue, we propose two strategies.

1. Support-weight median filter

After each warping step, we apply a support-weight median filter to the newly computed flow field. It further increases the robustness to reject outliers, because this filter considers the brightness similarity and geometric proximity between central pixels and neighboring pixels (Yoon and Kweon, 2006). Sun *et al.* (2010) quantitatively studied various aspects of the median filter to reduce noise for improving accuracy. Thus, we adopt the composited support-weight median filter proposed by Yoon and Kweon (2006) and Sun *et al.* (2010). This strategy is defined as

$$\min_{\bar{\boldsymbol{U}}_X} \sum_{Y\in N_X} L_{X,Y}|\bar{\boldsymbol{U}}_X - \boldsymbol{U}_Y|, \tag{13}$$

where $L_{X,Y} = \exp\big(-(\Delta a/\xi + \Delta b/\upsilon)r\big)$, $r = O(X)/O(Y)$, N_X denotes the set of neighbors of pixel X (it is set to 7×7 neighbors of the central pixel X in our experiments), $O(X)$ and $O(Y)$ are calculated using Eq. (22) in Sand and Teller (2008), $\boldsymbol{U}_Y$ is the displacement vector of the previous level, and $\bar{\boldsymbol{U}}_X$ is the initial displacement vector of the fine level. $\Delta a(X,Y)$ denotes the brightness similarity between $\boldsymbol{I}(X)$ and $\boldsymbol{I}(Y)$ measured as Euclidean distance and $\Delta b(X,Y)$ is the Euclidean distance, i.e., the proximity between X and Y. The parameters ξ and υ are used to tune the influence of brightness similarity and spatial proximity. Eq. (13) can be solved using the algorithm presented by Li and Osher (2009).

2. Downsampling strategy

We adopt the twice downsampling strategy to estimate the optical flow field. At the first stage, we use a five-level pyramid with a downsampling factor of 0.6. At the second stage, we use the result obtained from the first stage to initialize the flow, and a different factor of 0.85 is adopted in this pyramid downsampling. This can make up for the errors that are caused by using only a single downsampling factor. In experiments, we try multi-downsampling (for example, 3, 4, $\cdots$), and find that the results using twice downsampling are more accurate than those using single downsampling. Experiments show that the improvement using multi-sampling is not obvious as compared to twice downsampling. Concerning the computation time, we adopt the twice downsampling strategies here. Fig. 1 shows an example of the test images from the Middlebury dataset (Baker *et al.*, 2007) for optical flow estimation by single downsampling (with a downsampling factor of 0.85) and our twice downsampling with different coefficients. Figs. 1a and 1e are the test images of the 'Grove' and 'RubberWhale', respectively. Figs. 1b and 1f are their corresponding ground truth flows. Figs. 1c and 1g are the results using a single downsampling strategy. Figs. 1d and 1h are the results using our twice downsampling strategy with different coefficients. We use the corresponding average angular error (AAE) and the average end-point error (EPE) measures (Barron *et al.*, 1994). Our downsampling strategy gives smaller AAEs and EPEs than single downsampling for optical flow estimation, and the results are more accurate than those using single downsampling.

4 Experiments and evaluation

We implemented the algorithm in Matlab, and tested it on an Intel® Core™ Duo CPU T7300 2.00 GHz machine with 2 GB RAM. The experiments were performed on test image sequences of Middlebury datasets (Baker *et al.*, 2007). We empirically found the parameter values $\alpha = 0.05$, $\mu = 0.2$, $\lambda = 10^{-3}$, $\varepsilon = 10^{-6}$, $\xi = 90$, and $\upsilon = 90$ to be satisfactory, which we used in all experiments unless explicitly stated. It took 115 s for the 'Urban' image sequences at resolution 420×380.

4.1 Comparisons of different models

The first experiment shows the effectiveness of our data term. We compared the proposed data term with four different data terms in L^1 norm: (1) the brightness constancy constraint (BCC); (2) the gradient constancy constraint (GCC); (3) the combination of the brightness constancy and the gradient constancy constraints (BGCC) (equilibrium

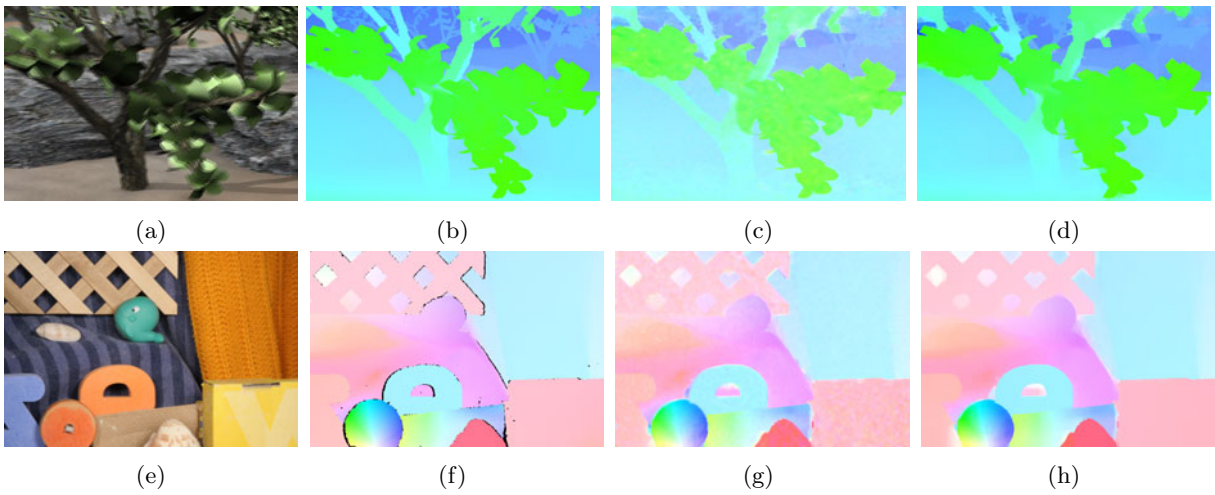


Fig. 1 The visual comparison of single downsampling and our twice downsampling. (a) 'Grove'; (b) The ground truth flow; (c) The result using single downsampling (AAE, 2.118°; EPE, 0.118); (d) The result using our twice downsampling (AAE, 1.306°; EPE, 0.089); (e) 'RubberWhale'; (f) The ground truth flow; (g) The result using single downsampling (AAE, 3.368°; EPE, 0.109); (h) The result using our twice downsampling (AAE, 2.634°; EPE, 0.084). AAE: average angular error; EPE: end-point error

parameter $\mu = 1$) (Here, our model also uses the equilibrium parameter $\mu = 1$); (4) the one using Kim *et al.* (2005)'s method. For fairness, we adopted the same regularization term to form the full model, and employed the uniform multi-resolution scheme to compute the optical flow field.

Fig. 2 shows the error comparison of the 'Grove2', 'Venus', 'Urban2', and 'Dimetrodon' sequences from Middlebury University (Baker *et al.*, 2007). We used the corresponding average angular error (AAE) measure (Barron *et al.*, 1994) to compare the quality of estimated flow fields with the ground truth. The AAEs for 'Urban2' and 'Venus' are small when using the BCC, and the GCC is favored in 'Grove2' and 'Dimetrodon' due to the illumination change. The AAEs using the BGCC are always between the BCC and the GCC. The AAEs using Kim's method are smaller than those using BCC, but at medium level. Since our data term adopts the brightness correction field combined with the gradient constraint, the results are more accurate than the others.

Fig. 3 shows the visual comparison for close-up of 'Grove3'. Image (a) is our whole estimated optical flow field from frame 10 and frame 11 of 'Grove3' in the Middlebury University database. Image (b) shows the close-up of the ground truth. Images (c) and (d) show close-up results by using BGCC and GCC, respectively. Image (e) is the result using Kim

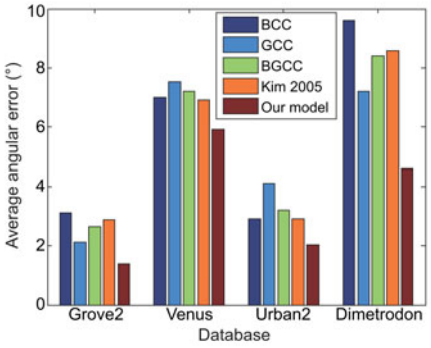


Fig. 2 Flow estimation average angular errors (AAE) with respect to different models. BCC: brightness constancy constraint; GCC: gradient constancy constraint; BGCC: brightness constancy and gradient constancy constraints

et al. (2005)'s method. As shown in the close-up images, some details of branches disappear when employing BGCC, GCC, and Kim's model. Our model is able to recover the branch details much better, as shown in image (f).

4.2 Optical flow benchmarking on Middlebury datasets

To evaluate the accuracy of current flow estimation techniques, benchmark sequences with a reliable ground truth are required. We used sequences of Middlebury datasets (Baker *et al.*, 2007) to test our method. Table 1 shows the AAEs of various

methods on different Middlebury sequences. Our method gives smaller AAEs than those methods for optical flow estimation.

Fig. 4 shows a comprehensive comparison of results produced by a number of approaches in the same framework. These methods contain: integrating the statistical framework with the linear model of the brightness variation (Kim *et al.*, 2005), using an appropriate adaptive regularization term (Wedel *et al.*, 2009b), employing an efficient median filter model (Sun *et al.*, 2010), and adopting an image-driven regularization term (Werlberger

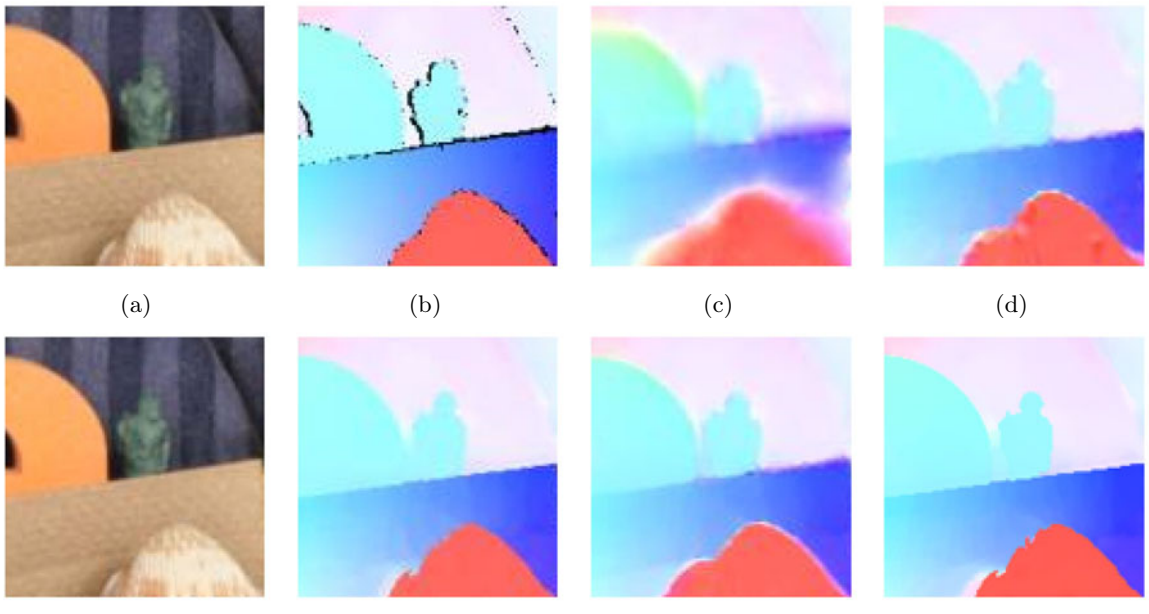
Table 1 Comparison of various methods regarding the average angular error (AAE) on different Middlebury sequences

MS	Average angular error (°)				
	Brox	Zach	Kim	Sun	Ours
RubberWhale	5.794	5.639	5.846	4.012	3.372
Hydrangea	5.496	5.536	5.385	2.873	1.965
Dimetrodon	5.297	4.610	4.575	2.136	2.273
Urban3	5.242	7.440	4.784	3.456	2.794
Grove3	6.356	7.471	5.845	4.987	4.768

MS: Middlebury sequence. Brox: Brox *et al.* (2004); Zach: Zach *et al.* (2007); Kim: Kim *et al.* (2005); Sun: Sun *et al.* (2010); Ours: our method

(a) (b) (c) (d) (e) (f)

Fig. 3 The visual comparison of the optical flow results. (a) Our estimated flow result; (b) The ground truth flow; (c) The close-up estimated flow result using the brightness and the gradient constraints (BGCC); (d) The close-up estimated flow result using the gradient constraint (GCC); (e) The close-up estimated flow result using Kim *et al.* (2005)'s method; (f) The close-up of our estimated flow result



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 4 The visual comparison of different methods. (a) Patch of frame 10; (b) The ground truth flow; (c) Integrating the statistical framework with the linear model of the brightness variation (Kim *et al.*, 2005); (d) Using an appropriate adaptive regularization term (Wedel *et al.*, 2009b); (e) Patch of frame 11; (f) Employing an efficient median filter model (Sun *et al.*, 2010); (g) Adopting an image-driven regularization term (Werlberger *et al.*, 2009); (h) Our result

et al., 2009). Using these methods, the edges of objects obtained are fuzzy, and the clearance between the wheel and the army men disappears. Our model produces an agreeable optical flow estimation in detail (Fig. 4h), due to the robust data term (integrating the brightness correction field and gradient constancy assumption) and the structure-adaptive regularization based on the Huber norm.

4.3 Complex-motion optical flow estimation

We compared the proposed method with a number of optical flow methods on some complex motion examples. Fig. 5 shows three examples containing complex articulated motion. The first row shows two-object-overlaid images from Middlebury datasets. From the second row to the fourth row, estimated optical flow results using the classical

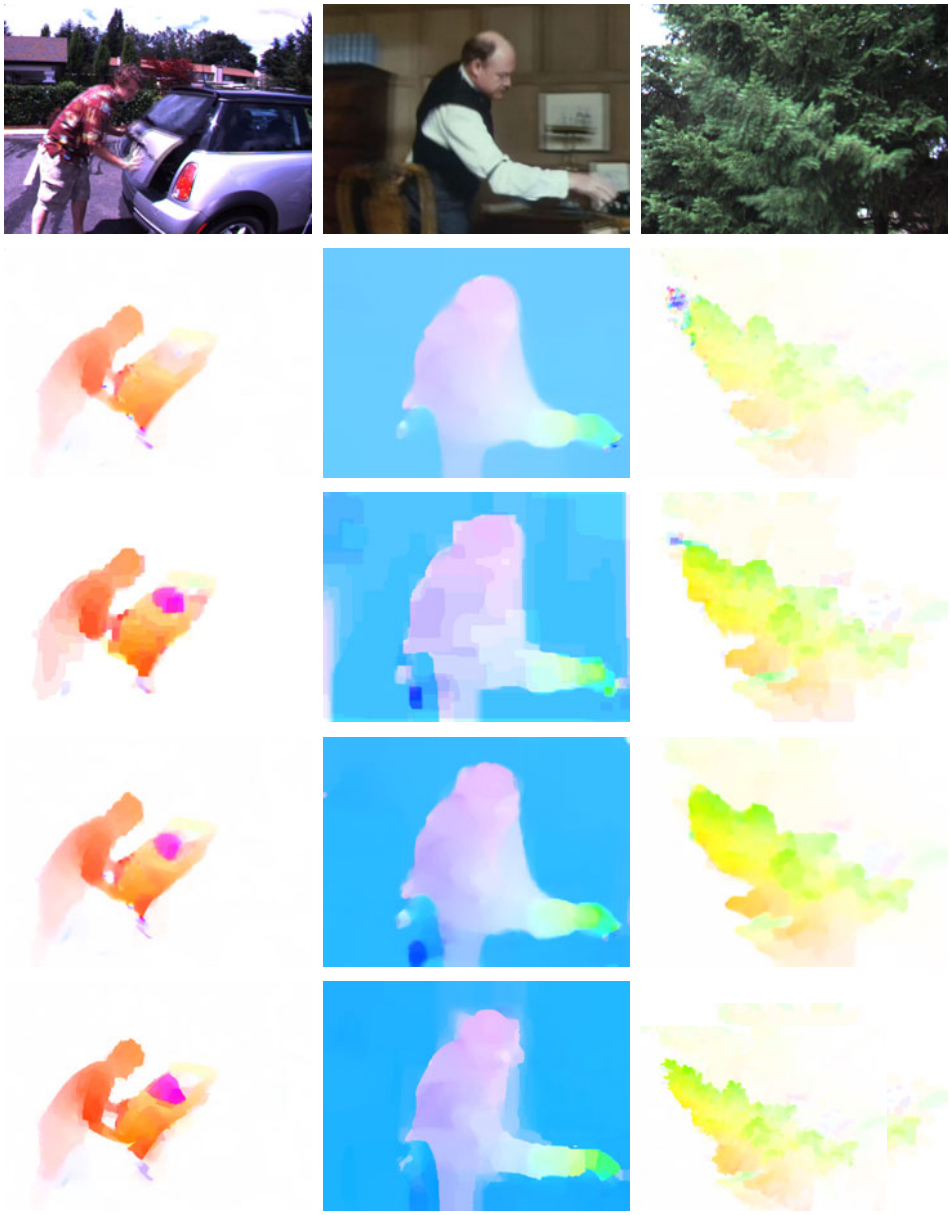


Fig. 5 The visual comparison on some complex motion examples. Top row: two-object-overlaid images from the Middlebury training dataset (Baker *et al.*, 2007); center top row: results obtained using the classic warping model (Brox *et al.*, 2004); center row: results obtained using SIFT flow (Liu *et al.*, 2008); center bottom row: results obtained using LDOF (Brox and Malik, 2011); bottom row: results obtained with our method

combination model (Brox *et al.*, 2004), large displacement optical flow (LDOF) (Brox and Malik, 2011), and scale-invariant feature transform (SIFT) flow (Liu *et al.*, 2008) are shown, respectively. The last row displays our estimated optical flow. There are some problems in motion details estimation using the classical combination model of Brox *et al.* (2004). For example, the motion of the rear windows in the left, the motion of elbow of the sword arm in the middle, and the motion of branches in the right are missing. The SIFT flow method shows typical discretization and quantization artifacts, in particular discontinuities aligned with the grid axes and block effects due to missing subpixel accuracy. Compared with SIFT flow, the results of LDOF are much better. The reason is that this method performs region-based descriptor matching and recovers large displacement flow by adjusting the objective to favor matching results. However, sometimes it is vulnerable to matching outliers, and does not estimate boundary details between regions well. Using the proposed model and the effective multi-resolution scheme, our experimental results are more accurate for estimating details of complex motion than the above methods, as shown in the last row of Fig. 5.

5 Conclusions and future work

In this paper, we present a new energy model for robust optical flow estimation. It consists of two terms: a data term combining a brightness correction field with a gradient constancy assumption, and a structure-adaptive regularization term based on the robust Huber norm. Compared with other methods, the proposed model can deal with slight changes of illumination conditions and reduces the piecewise constant of estimated flow fields in weakly textured areas from total variation. We solve the Euler-Lagrange equation of the proposed energy using a more effective multi-resolution method, which adopts the downsampling strategy twice with different downsampling factors. The proposed method recovers the missing motion details that are caused by using the traditional coarse-to-fine scheme. After each warping step, we apply a support-weight median filter to the newly computed flow field for rejecting outliers. The comparative experiments with state-of-the-art methods show that our method is more effective and produces more accurate results of optical flow estimation. Our future research will focus on estimating large disparities, which has a great promise on boosting the accuracy of motion estimation taking advantage of the proposed model without the necessity of using typical coarse-to-fine approaches.

Acknowledgements

We are grateful to the anonymous reviewers for their valuable comments and suggestions, and to Middlebury University for providing the experimental data.

References

Baker, S., Roth, S., Scharstein, D., Black, M., Lewis, J., Szeliski, R., 2007. A Database and Evaluation Methodology for Optical Flow. IEEE 11th Int. Conf. on Computer Vision, p.1-8. [doi:10.1109/ICCV.2007.4408903]

Barron, J., Fleet, D., Beauchemin, S., 1994. Performance of optical flow techniques. *Int. J. Comput. Vis.*, **12**(1): 43-77. [doi:10.1007/BF01420984]

Black, M.J., Anandan, P., 1996. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Comput. Vis. Image Understand.*, **63**(1):75-104. [doi:10.1006/cviu.1996.0006]

Brox, T., Malik, J., 2011. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, **33**(3):500-513. [doi:10.1109/TPAMI.2010.143]

Brox, T., Bruhn, A., Papenberg, N., Weickert, J., 2004. High Accuracy Optical Flow Estimation Based on a Theory for Warping. European Conf. on Computer Vision, p.25-36.

Bruhn, A., Weickert, J., Schnörr, C., 2005. Lucas/Kanade meets Horn/Schunk: combining local and global optical flow methods. *Int. J. Comput. Vis.*, **61**(3):211-231. [doi:10.1023/B:VISI.0000045324.43199.43]

Cassisa, C., Simoens, S., Prinet, V., 2009. Two-frame optical flow formulation in an unwarped multiresolution scheme. *LNCS*, **5856**:790-797. [doi:10.1007/978-3-642-10268-4_93]

Dessauer, M.P., Dua, S., 2010. Optical flow object detection, motion estimation, and tracking on moving vehicles using wavelet decompositions. *SPIE*, **7694**:76941J. [doi:10.1117/12.853281]

Efros, A., Berg, A., Mori, G., Malik, J., 2003. Recognizing Action at a Distance. Proc. 9th IEEE Int. Conf. on Computer Vision, p.726-733. [doi:10.1109/ICCV.2003.1238420]

Fakih, A., Zelek, J., 2008. Structure from Motion: Combining Features Correspondences and Optical Flow. 19th Int. Conf. on Pattern Recognition, p.1-4. [doi:10.1109/ICPR.2008.4761007]

Gennert, M.A., Negahdaripour, S., 1987. Relaxing the Brightness Constancy Assumption in Computing Optical Flow. Technical Report, Massachusetts Institute of Technology, Cambridge, MA, USA.

Gilland, D.R., Mair, B.A., Parker, J.G., 2008. Motion estimation for cardiac emission tomography by optical flow methods. *Phys. Med. Biol.*, **53**(11):2991-3006. [doi:10.1088/0031-9155/53/11/016]

Gray, R.M., 2006. Toeplitz and circulant matrices: a review foundations and trends. *Commun. Inform. Theory*, **2**(3):155-239. [doi:10.1561/0100000006]

Haussecker, H.W., Fleet, D.J., 2001. Computing optical flow with physical models of brightness variation. *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**(6):661-673. [doi:10.1109/34.927465]

Horn, B., Schunck, B., 1981. Determining optical flow. *Artif. Intell.*, **17**(1-3):185-203. [doi:10.1016/0004-3702(81)90024-2]

Hsiao, I., Rangarajan, A., Gindi, G., 2003. A new convex edge-preserving median prior with applications to tomography. *IEEE Trans. Med. Imag.*, **22**(5):580-585. [doi:10.1109/TMI.2003.812249]

Huber, P.J., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Stat.*, **1**(5):799-821. [doi:10.1214/aos/1176342503]

Kim, Y.H., Martinez, A.M., Kak, C.A., 2005. Robust motion estimation under varying illumination. *Image Vis. Comput.*, **23**(4):365-375. [doi:10.1016/j.imavis.2004.05.010]

Lempitsky, V., Roth, S., Rother, C., 2008. FusionFlow: Discrete Continuous Optimization for Optical Flow Estimation. IEEE Conf. on Computer Vision and Pattern Recognition, p.1-8. [doi:10.1109/CVPR.2008.4587751]

Li, Y., Osher, S., 2009. A new median formula with applications to PDE based denoising. *Commun. Math. Sci.*, **7**(3):741-753.

Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T., 2008. SIFT flow: dense correspondence across different scenes. *LNCS*, **5304**:28-42. [doi:10.1007/978-3-540-88690-7_3]

Myronenko, A., Song, X., 2009. Image Registration by Minimization of Residual Complexity. IEEE Conf. on Computer Vision and Pattern Recognition, p.49-56. [doi:10.1109/CVPR.2009.5206571]

Negahdaripour, S., 1998. Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, **20**(9):961-979. [doi:10.1109/34.713362]

Sand, P., Teller, S., 2008. Particle video: long-range motion estimation using point trajectories. *Int. J. Comput. Vis.*, **80**(1):72-91. [doi:10.1007/s11263-008-0136-6]

Shulman, D., Herve, J.Y., 1989. Regularization of Discontinuous Flow Fields. Proc. Workshop on Visual Motion, p.81-86. [doi:10.1109/WVM.1989.47097]

Steinbrücker, F., Pock., T., 2009. Large Displacement Optical Flow Computation without Warping. Int. Conf. on Computer Vision, p.1609-1614.

Strang, G., 1999. The discrete cosine transform. *SIAM Rev.*, **41**(1):135-147. [doi:10.1137/S0036144598336745]

Sun, D., Roth, S., Black, M.J., 2010. Secrets of Optical Flow Estimation and Their Principles. IEEE Conf. on Computer Vision and Pattern Recognition, p.2432-2439. [doi:10.1109/CVPR.2010.5539939]

Teng, C.H., Lai, S.H., Chen, Y.S., Hsu, W.H., 2005. Accurate optical flow computation under non-uniform brightness variations. *Comput. Vis. Image Understand.*, **97**(3):315-346. [doi:10.1016/j.cviu.2004.08.002]

Wang, M.Y., Hu, H.B., Qin, B.J., 2007. Robust Deformable Medical Image Registration Using Optical Flow and Multilevel Free Form Deformation. Nuclear Science Symp. Conf. Record, p.4552-4555.

Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D., 2009a. An improved algorithm for TV-L^1 optical flow. *LNCS*, **5604**:23-45. [doi:10.1007/978-3-642-03061-1_2]

Wedel, A., Cremers, D., Pock, T., Bischof, H., 2009b. Structure- and Motion-Adaptive Regularization for High Accuracy Optic Flow. IEEE 12th Int. Conf. on Computer Vision, p.1663-1668. [doi:10.1109/ICCV.2009.5459375]

Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., Bischof, H., 2009. Anisotropic Huber-L^1 Optical Flow. British Machine Vision Conf., p.1-11.

Yoon, K.J., Kweon, I.S., 2006. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.*, **28**(4):650-656. [doi:10.1109/TPAMI.2006.70]

Zach, C., Pock, T., Bischof, H., 2007. A duality based approach for realtime TV-L^1 optical flow. *LNCS*, **4713**:214-223. [doi:10.1007/978-3-540-74936-3_22]

Appendix: derivation of Eq. (12)

We rewrite Eq. (11) as follows:

$$\begin{aligned}E(\boldsymbol{U}) =& \sum_{n=1}^{N}\log\left((\boldsymbol{d}_n^{\mathrm{T}}\boldsymbol{r})^2/\alpha+1\right)\\&+\mu\sum_{\boldsymbol{X}}||\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})-\nabla\boldsymbol{I}_1(\boldsymbol{X})||_1\\&+\lambda\sum_{\boldsymbol{X}}||w(\boldsymbol{X})\nabla\boldsymbol{U}||_\delta.\end{aligned}\tag{A1}$$

For convenience, we assume $E(\boldsymbol{U}) = E_1(\boldsymbol{U}) + E_2(\boldsymbol{U})$.

$$E_1(\boldsymbol{U}) = \sum_{n=1}^{N}\log\left((\boldsymbol{d}_n^{\mathrm{T}}\boldsymbol{r})^2/\alpha+1\right),\tag{A2}$$

$$\begin{aligned}E_2(\boldsymbol{U}) =&\mu\sum_{\boldsymbol{X}}||\nabla\boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U})-\nabla\boldsymbol{I}_1(\boldsymbol{X})||_1\\&+\lambda\sum_{\boldsymbol{X}}||\nabla\boldsymbol{U}||_\delta.\end{aligned}\tag{A3}$$

According to the analysis of Myronenko and Song (2009), $\boldsymbol{d}_n^{\mathrm{T}}\boldsymbol{r}$ is the discrete cosine transform of $\boldsymbol{r}$, and it denotes $\boldsymbol{A} = \mathrm{dct}\{\boldsymbol{r}\}$. The dct and idct are discrete cosine transform and inverse discrete cosine transform, respectively.

$$E_1(\boldsymbol{U}) = \sum_{n=1}^{N} \log(\boldsymbol{A}^2/\alpha + 1). \tag{A4}$$

According to the calculus of variations, a minimizer of Eq. (A1) must fulfill the Euler-Lagrange equation:

$$\delta E_1 + \delta E_2 = 0, \tag{A5}$$

where δ is the symbol of variation. Thus, through the calculus of variations, we obtain

$$\delta E_1 = -\operatorname{idct}\left(\frac{2\boldsymbol{A}/\alpha}{\boldsymbol{A}^2/\alpha + 1}\right), \tag{A6}$$

$$\begin{aligned}\delta E_2 =& \mu\nabla\varphi(||\nabla \boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U}) - \nabla \boldsymbol{I}_1(\boldsymbol{X})||_1^2) \\ & - \lambda \operatorname{div}(\phi^{'}(\nabla \boldsymbol{U})\nabla \boldsymbol{U}),\end{aligned} \tag{A7}$$

where $\varphi(x^2) = \sqrt{x^2+\varepsilon}$, $\varepsilon > 0$, $\phi(\nabla \boldsymbol{U}) = ||w(\boldsymbol{X})\nabla \boldsymbol{U}||_\delta$, $\delta > 0$. Finally, Eq. (A8) is obtained:

$$\begin{aligned}&-\operatorname{idct}\left(\frac{2\boldsymbol{A}}{\boldsymbol{A}^2+\alpha}\right)\nabla \boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U}) \\ &\quad + \mu\nabla\varphi(||\nabla \boldsymbol{I}_2(\boldsymbol{X}+\boldsymbol{U}) - \nabla \boldsymbol{I}_1(\boldsymbol{X})||_1^2) \\ &\quad - \lambda \operatorname{div}(\phi'(\nabla \boldsymbol{U})\nabla \boldsymbol{U}) = 0.\end{aligned} \tag{A8}$$