

Robust Optical Flow Estimation of Double-Layer Images under Transparency or Reflection

Jiaolong Yang*
BIT & ANU

Hongdong Li
ANU & ACRV

Yuchao Dai
CECS, ANU

Robby T. Tan
Yale-NUS College

Abstract

This paper deals with a challenging, frequently encountered, yet not properly investigated problem in two-frame optical flow estimation. That is, the input frames are compounds of two imaging layers – one desired background layer of the scene, and one distracting, possibly moving layer due to transparency or reflection. In this situation, the conventional brightness constancy constraint – the cornerstone of most existing optical flow methods – will no longer be valid. In this paper, we propose a robust solution to this problem. The proposed method performs both optical flow estimation, and image layer separation. It exploits a generalized double-layer brightness consistency constraint connecting these two tasks, and utilizes the priors for both of them. Experiments on both synthetic data and real images have confirmed the efficacy of the proposed method. To the best of our knowledge, this is the first attempt towards handling generic optical flow fields of two-frame images containing transparency or reflection.

1. Introduction

Most optical flow methods assume that there is only one imaging layer on the observed image with the brightness of scene objects, and use the brightness constancy constraint (BCC) to estimate the optical flow for scene objects. This single imaging layer assumption, however, can be often violated in real-world situations, especially in cases involving transparency or reflection. Transparencies and reflections are frequently met in imaging process, *e.g.*, when one is looking at street scene from inside a car through a stained windscreen, or seeing through a thin layer of rain, looking into a window with semi-reflections on the window surface *etc.* The BCC will generally not hold for the resultant double-layer images, even in ideal noise-free cases.

In all the above examples, the observed image \mathbf{I} can be

*J. Yang is with Beijing Lab of Intelligent Information Technology, Beijing Institute of Technology (BIT), and with College of Engineering and Computer Science, the Australian National University (ANU). Email: jiaolong.yang@anu.edu.au

modeled as a superposition of two constituting layers, denoted as $\mathbf{I} = \mathbf{L}_1 \oplus \mathbf{L}_2$, where \oplus denotes some suitable layer combination operator. Without loss of generality, we call \mathbf{L}_1 the background scene layer, which corresponds to the image of the desired scene that we intend to capture, and \mathbf{L}_2 the foreground distracting layer, which corresponds to the semi-transparent media (*e.g.* a glass window with dirt or reflections on it) or the semi-reflected image.

The main goal of this paper is to robustly estimate the optical flow field of the scene objects (*i.e.* the background layer), which is of concern for vision systems. We consider two general cases: the foreground distracting layer is *stationary*, or *dynamically changing*.

Let \mathbf{I} and \mathbf{I}' be two time-consecutive frames of a scene containing the aforementioned two layers. In the presence of a dynamic foreground layer, there are two legitimate optical flow fields – one for the foreground layer and one for the background layer. Denote the two flow fields generated by the movements of the two layers as \mathbf{U} and \mathbf{V} respectively. The relationships among the observed images, the image layers and the optical flow fields can be given as

$$\begin{array}{ccc} \mathbf{I} & = & \mathbf{L}_1 \oplus \mathbf{L}_2 \\ & & \downarrow \mathbf{U} \quad \downarrow \mathbf{V} \\ \mathbf{I}' & = & \mathbf{L}'_1 \oplus \mathbf{L}'_2 \end{array}$$

When $\mathbf{V} \equiv \mathbf{0}$ and $\mathbf{L}_2 \equiv \mathbf{L}'_2$, *i.e.* the foreground layer is static, our task is to estimate a single flow field \mathbf{U} for background layer, and also estimate the layers \mathbf{L}_1 , \mathbf{L}'_1 , \mathbf{L}_2 . Otherwise when a dynamic foreground layer exists, we will estimate two flow fields \mathbf{U} , \mathbf{V} as well as the layers \mathbf{L}_1 , \mathbf{L}'_1 , \mathbf{L}_2 , \mathbf{L}'_2 . As we explicitly perform image layer separation (*i.e.* estimating \mathbf{L}_1 , \mathbf{L}'_1 , \mathbf{L}_2 , \mathbf{L}'_2), an appealing byproduct of our method is the restoration of the clear scene images.

For either of the two cases with a static or dynamic foreground layer, this is a highly ill-posed problem, especially considering optical flow estimation and image layer separation problems *per se* are known to be ill-posed. From only two input images, our task is to recover one or two optical fields, as well as the two unknown layers.

Little work has been reported in the literature concerning

this double-layer image optical flow estimation problem, with only a few exceptions in the early days of computer vision research, *e.g.* [28][29][18][11]. These works however often used over-simplified assumption and restrictive motion field models, such as assuming a constant flow field over time or over space (*e.g.* globally translating). Bergen *et al.* [3] proposed a “three-frame algorithm” to recover two constituting flow fields, assuming the flow field is constant over at least three frames.

In contrast, this paper removes these restrictive assumptions, and proposes a two-frame algorithm for robustly recovering the flow field(s). Our method works for generic motions, and is thus applicable to a much wider range of practical situations for robust optical flow estimation.

1.1. Related Works

This paper is concerned with optical flow estimation in double-layer images where both layers can possibly be moving. Despite that the phenomena of such multiple imaging layers and motions are frequently encountered in reality, few papers in the literature have been devoted to this topic. This is in a sharp contrast to the existence of vast amount of papers on the classic optical flow problems (an analysis of recent practices of optical flow can be found in [31]).

One of the first work for multiple optical flow computation is possibly due to Shiwaza *et al.* [28][29]. By assuming the two underlying flow fields to be constant (*e.g.* pure translating), they derived a generalized brightness constancy constraint for the multi-motion case. However, this constant motion assumption is restrictive, not applicable for general flow fields with complex motions. Nevertheless, their method, being one of the first, has inspired a number of variants and extensions [24][1][25][36]. Some variants operate in the Fourier domain, *e.g.* [17][18][11].

The flow estimation problem for two-layer images in this paper should not be confused with those works concerning “motion-layer segmentation”, albeit the two do share some similarity and the boundary between them can sometimes be fuzzy. For example, Wang and Adelson [38] proposed to segment the image layers based on a pre-computed optical flow field. Irani *et al.* [15] used temporal integration to track occluding or transparent moving objects with parametric motion. Black and others [4][16][32][42] proposed a number of algorithms for multiple parametric motion estimation and segmentation. Yang and Li [44] fit a flow field with piecewise parametric models. Weiss [39] presented a nonparametric motion estimation and segmentation method to handle generic smooth motions, thus this method is more related to ours. However, the method of Weiss and most other aforementioned methods primarily focused on image and motion segmentation, while we decompose the whole image into two composite brightness layers, and compute one generic flow field on each layer.

The proposed method involves solving two tasks simultaneously: optical flow field estimation, and reflection/transparent layer separation. For the second task, many researches have been published previously. For example, Levin *et al.* [20][19] proposed methods for separating an image into two transparent layers using local statistics priors of natural images. Single image solutions are also investigated in [22] and [46]. To utilize multiple frames, layer separation methods have been proposed based on aligning the frames with one layer [41][21][14] or multiple layers [33][13]. Sarel and Irani [26] presented an information theory based approach for separating transparent layers by minimizing the correlation between the layers. Chen *et al.* [10] gave a gradient domain approach for moving layer separation which is also based on information theory. Schechner *et al.* [27] developed a method for layer separation using image focus as a cue. By using independent component analysis, Farid and Adelson [12] proposed a layer separation method which works on multiple observations under different mixing weights. Techniques for image layer separation were also developed in the field of intrinsic image/video extraction [35][40][45].

In the context of stereo matching with transparency, Szeliski and Golland [34] simultaneously recovered disparities, true colors, and opacity of visible surface elements. Tsing *et al.* [37] estimated both depth and colors of the component layers. Li *et al.* [23] proposed a simultaneous video defogging and stereo matching algorithm.

The recent work of Xue *et al.* [43] has a very similar formulation compared to ours. However, the goal and motivation of obstruction-free photography from a video sequence in [43] are different from ours. The underlying assumptions on the flow fields, the employed flow solvers and the initialization techniques are dissimilar.

2. Problem Setup

For ease of presentation, in formulating the problem (Sec. 2 and Sec. 3) and presenting the optimization (Sec. 4), we will focus on the dynamic foreground case (*i.e.* double-layer flow estimation). The static foreground case (*i.e.* single-layer flow estimation) is simpler and can be derived accordingly. Note that, the static foreground case, though relatively simpler, is also of interest and very challenging.

2.1. Linear Additive Imaging Model

In previous discussion, we simply used $\mathbf{I} = \mathbf{L}_1 \oplus \mathbf{L}_2$ to denote the layer superposition operation, but did not give its exact form. To make the idea of this paper more concrete, we opt for the linear additive model $+$ as a concrete example for \oplus , *i.e.*, $\mathbf{I} = \mathbf{L}_1 + \mathbf{L}_2$.

The linear additive model itself, while simple, has been used successfully in the past in solving many vision problems involving transparency and reflection (*e.g.*, in shadow

removal [46], image matting [34] and reflection separation [22]). Moreover, by applying logarithm operation, a multiplicative superposition model can also be converted to an additive one.

Taking two frames of observations, \mathbf{I} and \mathbf{I}' , at two consecutive time steps t and $t + 1$, we have

$$\mathbf{I}(\mathbf{X}) = \mathbf{L}_1(\mathbf{X}) + \mathbf{L}_2(\mathbf{X}), \quad (1)$$

$$\mathbf{I}'(\mathbf{X}) = \mathbf{L}'_1(\mathbf{X}) + \mathbf{L}'_2(\mathbf{X}), \quad (2)$$

where \mathbf{X} is a matrix indexing all pixel coordinates.

2.2. Double Layer Brightness Constancy

In the presence of transparencies or reflections, it is important to note that the conventional BCC condition cannot be applied directly to the observed images. Below, we will derive a generalized BCC condition which is applicable to the double-layer case.

The basic assumption that we will base our method on is: any component layer of the observed image must satisfy the brightness constancy condition individually. This is a realistic and mild assumption which is applicable to a wide range of transparency and reflection phenomena encountered in natural images. Cases that violate this basic assumption are deemed beyond the scope of this current paper.

Suppose, during two small time steps, layer \mathbf{L}_1 changed to \mathbf{L}'_1 according to a motion field of \mathbf{U} , and layer \mathbf{L}_2 changed to \mathbf{L}'_2 according to a different motion field \mathbf{V} . Based on the assumption that the brightness of the objects in each individual layer is constant, we have

$$\mathbf{L}_1(\mathbf{X}) = \mathbf{L}'_1(\mathbf{X} + \mathbf{U}), \quad (3)$$

$$\mathbf{L}_2(\mathbf{X}) = \mathbf{L}'_2(\mathbf{X} + \mathbf{V}). \quad (4)$$

Together with the imaging model in (1) and (2), we call the above constraints the *generalized double-layer BCC condition* for an input double-layer image pair $(\mathbf{I}, \mathbf{I}')$.

2.3. The Double Layer Optical Flow Problem

Given the above linear additive imaging model as well as the generalized BCC conditions, we aim to recover both $\mathbf{L}_1, \mathbf{L}'_1, \mathbf{L}_2, \mathbf{L}'_2$ and \mathbf{U}, \mathbf{V} .

To make this severely ill-posed problem trackable, we adopt the energy minimization framework, and base it on the generalized BCC conditions as well as priors for optical flows and image layers. The energy function reads as

$$E = E_B + \lambda_L E_L + \lambda_F E_F, \quad (5)$$

where E_B corresponds to the double-layer BCC condition, E_L and E_F are the regularization terms (or prior terms) for the latent image layers, and the unknown optical flow fields, respectively. The λ s are trade-off parameters.

In energy (5), we use $E_B = E_B(\mathbf{L}_1, \mathbf{L}'_1, \mathbf{L}_2, \mathbf{L}'_2, \mathbf{U}, \mathbf{V})$ to represent the BCC condition in the following way¹:

$$E_B = \|\mathbf{L}_1(\mathbf{X}) - \mathbf{L}'_1(\mathbf{X} + \mathbf{U})\| + \|\mathbf{L}_2(\mathbf{X}) - \mathbf{L}'_2(\mathbf{X} + \mathbf{V})\|. \quad (6)$$

We use $\|\cdot\|$ to denote the ℓ_1 -norm in this paper unless otherwise specified. We choose to use ℓ_1 -norm as the cost function mainly for its robustness [6, 47] and its convenience in optimization. The two regularization terms E_L and E_F will be detailed in the following section.

3. Regularization

Using prior information as regularization is a common practice for solving ill-posed problems. In this paper, the task is to separate the input frames into latent layers, and to recover the associated flow fields.

Priors are generally task-dependent. By enforcing different priors to latent layers and to optical flow fields, the algorithm can be adapted to solving different tasks. For example, if one knows the two latent layers are images of natural scenes, then the layers can be assumed to have sparse gradients (*i.e.*, satisfying the well-known natural image priors). Moreover, for general optical flow fields, one can assume they are piecewise constant or piecewise smooth.

3.1. Natural Image Prior: Sparse Gradient

The research in natural image statistics shows that images of typical real-world scenes obey sparse spatial gradient distributions [35, 19]. The distribution of a natural image \mathbf{L} can often be modeled as a generalized Laplace distribution (*a.k.a.*, generalized Gaussian distribution), *i.e.*,

$$P(\mathbf{L}) = \prod_{\mathbf{x} \in \mathbf{X}} \exp(-|\partial_x \mathbf{L}(\mathbf{x})|^p - |\partial_y \mathbf{L}(\mathbf{x})|^p), \quad (7)$$

where the power p is a parameter usually within $[0.0, 1.0]$. A convenient choice is $p = 1$, with which the energy is reduced to the ℓ_1 -norm of image spatial gradients. For ease exposition, we will let $p = 1$ in this paper, though bear in mind that using other values of p is possible and may be advantageous in particular applications.

Taking the negative logarithm, the prior in (7) can be represented in the energy minimization form, *i.e.*

$$\|\nabla \mathbf{L}(\mathbf{X})\| \rightarrow \min, \quad (8)$$

where $\nabla = (\partial_x, \partial_y)^\top$. Therefore, the latent layer regularization term $E_L = E_L(\mathbf{L}_1, \mathbf{L}'_1, \mathbf{L}_2, \mathbf{L}'_2)$ can be written as

$$E_L = \|\nabla \mathbf{L}_1(\mathbf{X})\| + \|\nabla \mathbf{L}'_1(\mathbf{X})\| + \|\nabla \mathbf{L}_2(\mathbf{X})\| + \|\nabla \mathbf{L}'_2(\mathbf{X})\|. \quad (9)$$

¹For brevity, hereafter we use a short-hand notation for functions defined on all pixel coordinates \mathbf{X} : a function $f(\mathbf{X})$ should be understood as $\sum_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$, unless otherwise specified.

3.2. Optical Flow Priors: Spatial Smoothness

Early methods for solving multi-layer optical flow problem often made restrictive assumption about the unknown flow fields. For example, [3] proposed a three-frame algorithm for recovering two component motion fields by assuming that the motion fields are constant over time, and [28] was built upon a local constant motion assumption to derive its basic equation. In this paper, these restrictions are removed and the proposed method can handle more general and more complex motion fields.

We use a general assumption on flow field, namely, the optical flows are generally piecewise constant or piecewise smooth. To capture this prior, we adopt the total variation (TV) model [47] or total generalized variation (TGV) model [5]. Specifically, a flow field \mathbf{U} will be regularized by the following energy:

$$\|\mathbf{U}\|_{\text{TGV}^k} \rightarrow \min, \quad (10)$$

where $\|\mathbf{U}\|_{\text{TGV}^k} \doteq \text{TGV}^k(\mathbf{U}_x) + \text{TGV}^k(\mathbf{U}_y)$, and $\text{TGV}^k(\cdot)$ denotes the k -th order TGV measure for horizontal and vertical flow components \mathbf{U}_x and \mathbf{U}_y .

In general, the k -th order TGV favors solutions that are piecewise composed of $(k-1)$ -th order polynomials: with $k=1$, TGV^1 reduces to the TV model which favors piecewise constant fields; with $k=2$, TGV^2 favors piecewise affine fields. We will only consider TV and TGV^2 in this paper, and the resultant prior regularization term $E_F = E_F(\mathbf{U}, \mathbf{V})$ for the flow fields can be written as

$$E_F(\mathbf{U}, \mathbf{V}) = \|\mathbf{U}\|_{\text{TGV}^k} + \|\mathbf{V}\|_{\text{TGV}^k}. \quad (11)$$

where $k=1$ (*i.e.* TV) or 2.

4. Energy Minimization

4.1. The Overall Objective Function

By stacking all the constraints over both latent layers and flow fields, we reach an energy minimization problem as

$$\begin{aligned} \min E(\mathbf{L}_1, \mathbf{L}'_1, \mathbf{L}_2, \mathbf{L}'_2, \mathbf{U}, \mathbf{V}) &= E_B + \lambda_L E_L + \lambda_F E_F \\ &= (\|\mathbf{L}_1(\mathbf{X}) - \mathbf{L}'_1(\mathbf{X} + \mathbf{U})\| + \|\mathbf{L}_2(\mathbf{X}) - \mathbf{L}'_2(\mathbf{X} + \mathbf{V})\|) \\ &\quad + \lambda_L (\|\nabla \mathbf{L}_1\| + \|\nabla \mathbf{L}'_1\| + \|\nabla \mathbf{L}_2\| + \|\nabla \mathbf{L}'_2\|) \\ &\quad + \lambda_F (\|\mathbf{U}\|_{\text{TGV}^k} + \|\mathbf{V}\|_{\text{TGV}^k}), \end{aligned} \quad (12)$$

subject to

$$\mathbf{I} = \mathbf{L}_1 + \mathbf{L}_2, \quad \mathbf{I}' = \mathbf{L}'_1 + \mathbf{L}'_2, \quad (13)$$

$$\mathbf{0} \leq \mathbf{L}_2 \leq \min(\mathbf{I}, c), \quad \mathbf{0} \leq \mathbf{L}'_2 \leq \min(\mathbf{I}', c). \quad (14)$$

where the \mathbf{X} 's in the gradient terms of (9) are omitted for brevity.

Note that, to distinguish background and foreground layers, we introduce in (14) the element-wise bound constraints on the layers. We assume the foreground layer containing transparency or reflection has weaker signal, and use

a small constant scalar c (*e.g.* $c=0.25$ for brightness values in the range of $[0,1]$) as its brightness upper bound. This can be understood as an additional bound prior for layer separation. Also note that, putting aside (14), there is a global shift ambiguity for the layer values: adding an arbitrary scalar $s \in \mathbb{R}$ to $\mathbf{L}_1, \mathbf{L}'_1$ then $-s$ to $\mathbf{L}_2, \mathbf{L}'_2$ does not change the energy in (12), nor does it affect (13). This is because all the terms in (12) depend on value difference rather than absolute value. Nevertheless, both the lower and upper bounds in (14) help constrain the absolute values.

4.2. Alternated Minimization

To solve the above energy minimization problem, we first substitute the additive model constraints in (13) as hard constraints to eliminate \mathbf{L}_1 and \mathbf{L}'_1 in (12). Consequently, the energy function is now defined only on latent layers $\mathbf{L}_2, \mathbf{L}'_2$ and optical flows \mathbf{U}, \mathbf{V} .

Then, examining the energy form in (12), we notice that: *i)* the prior terms for optical flow field, *i.e.* E_F , is independent of the prior term for latent layers E_L ; and *ii)* the BCC energy term E_B is the only term that links the flow estimation with latent layer separation. Based on these observations, we solve the minimization problem via block coordinate descent in an alternating fashion.

Specifically, starting from a proper initialization, our algorithm alternately solves the following two sub-problems:

- **(Layer Separation):** Given current flow field estimates $\{\mathbf{U}, \mathbf{V}\}$, solve for image layers $\{\mathbf{L}_2, \mathbf{L}'_2\}$ via the following minimization:

$$\min_{\mathbf{L}_2, \mathbf{L}'_2} (E_B(\mathbf{L}_2, \mathbf{L}'_2) + \lambda_L E_L(\mathbf{L}_2, \mathbf{L}'_2)). \quad (15)$$

- **(Flow Computation):** Given current image layers $\{\mathbf{L}_2, \mathbf{L}'_2\}$, estimate $\{\mathbf{U}, \mathbf{V}\}$ by solving the following two-layer optical flow problem:

$$\min_{\mathbf{U}, \mathbf{V}} (E_B(\mathbf{U}, \mathbf{V}) + \lambda_F E_F(\mathbf{U}, \mathbf{V})). \quad (16)$$

More details are given below.

4.2.1 Update the image layers

Given current optical flow estimates \mathbf{U} and \mathbf{V} , the latent image layers $\mathbf{L}_2, \mathbf{L}'_2$ can be updated by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{L}_2, \mathbf{L}'_2} &\|(\mathbf{I} - \mathbf{L}_2)(\mathbf{X}) - (\mathbf{I}' - \mathbf{L}'_2)(\mathbf{X} + \mathbf{U})\| + \|\mathbf{L}_2(\mathbf{X}) - \mathbf{L}'_2(\mathbf{X} + \mathbf{V})\| \\ &+ \lambda_L (\|\nabla(\mathbf{I} - \mathbf{L}_2)\| + \|\nabla(\mathbf{I}' - \mathbf{L}'_2)\| + \|\nabla \mathbf{L}_2\| + \|\nabla \mathbf{L}'_2\|) \\ \text{subject to } &\mathbf{0} \leq \mathbf{L}_2 \leq \min(\mathbf{I}, c), \quad \mathbf{0} \leq \mathbf{L}'_2 \leq \min(\mathbf{I}', c), \end{aligned} \quad (17)$$

This is a convex optimization problem defined on \mathbf{L}_2 and \mathbf{L}'_2 , and the cost function can be arranged into

$$\begin{aligned} \min_{\mathbf{l}} &\|\mathbf{A} \cdot \mathbf{l} - \mathbf{b}\|, \\ \text{subject to } &lb_i \leq l_i \leq ub_i, \forall i \end{aligned} \quad (18)$$

where \mathbf{A} and \mathbf{b} encode all the ℓ_1 constraints on latent layers, which are extremely sparse (only a few elements in each row are non-zero). \mathbf{I} is a column vector containing elements in \mathbf{L}_2 and \mathbf{L}'_2 . lb_i and ub_i are constant bounds from (14). The constraints are linear function of the latent layers \mathbf{L}_2 and \mathbf{L}'_2 , thus this problem can be solved as a linear programming using off-the-self solvers.

Nevertheless, to utilize the sparse structure in the problem and speed up the implementation, we solve the problem by using a tailored version of Iteratively Reweighted Least Squares (IRLS) [9]. With IRLS, one can also adapt the formulation to different priors readily, e.g., replacing ℓ_1 -norm with ℓ_p -norm ($0 < p < 1$). Details of our IRLS variant can be found in the *Supplementary Material*.

Use of color images. The above formulations can be easily extended to color RGB images. With color images, the double-layer BCC term E_B and layer regularization term E_L will be evaluated at R-G-B channels separately. The flow fields \mathbf{U} and \mathbf{V} are shared by all three channels.

4.2.2 Update the flow fields \mathbf{U} and \mathbf{V}

Given current layer estimates $\mathbf{L}_2, \mathbf{L}'_2$, and $\mathbf{L}_1 = \mathbf{I} - \mathbf{L}_2, \mathbf{L}'_1 = \mathbf{I}' - \mathbf{L}'_2$, the next step is to update the associated two flow fields \mathbf{U} and \mathbf{V} . This is done by solving the following optimization problem:

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{L}_1(\mathbf{X}) - \mathbf{L}'_1(\mathbf{X} + \mathbf{U})\| + \|\mathbf{L}_2(\mathbf{X}) - \mathbf{L}'_2(\mathbf{X} + \mathbf{V})\| + \lambda_F (\|\mathbf{U}\|_{\text{TGV}^k} + \|\mathbf{V}\|_{\text{TGV}^k}). \quad (19)$$

The computations for these two flow fields are in fact separable. This can be easily seen from the above optimization, as the cost function can be expressed as the sum of two terms, each of which can be solved in isolation, i.e., given $\{\mathbf{L}_1, \mathbf{L}'_1, \mathbf{L}_2, \mathbf{L}'_2\}$, solve

$$\min_{\mathbf{U}} \|\mathbf{L}_1(\mathbf{X}) - \mathbf{L}'_1(\mathbf{X} + \mathbf{U})\| + \lambda_F \|\mathbf{U}\|_{\text{TGV}^k}, \quad (20)$$

$$\min_{\mathbf{V}} \|\mathbf{L}_2(\mathbf{X}) - \mathbf{L}'_2(\mathbf{X} + \mathbf{V})\| + \lambda_F \|\mathbf{V}\|_{\text{TGV}^k}. \quad (21)$$

To solve the above optical flow problems, we use quadratic relaxation and introduce an auxiliary flow field to decouple the BCC term and regularization term, similar to [47, 30]. To solve the resulting TV- L_2 (a.k.a. the ROF model) and TGV^2 - L_2 problem, we apply the primal-dual method of [8] which is GPU-friendly. Details of our algorithm and implementation can be found in the *Supplementary Material*.

5. Experimental Results

In this section, we validate the proposed model and framework, and evaluate the performance of our method. We report the experimental results on both synthetic data

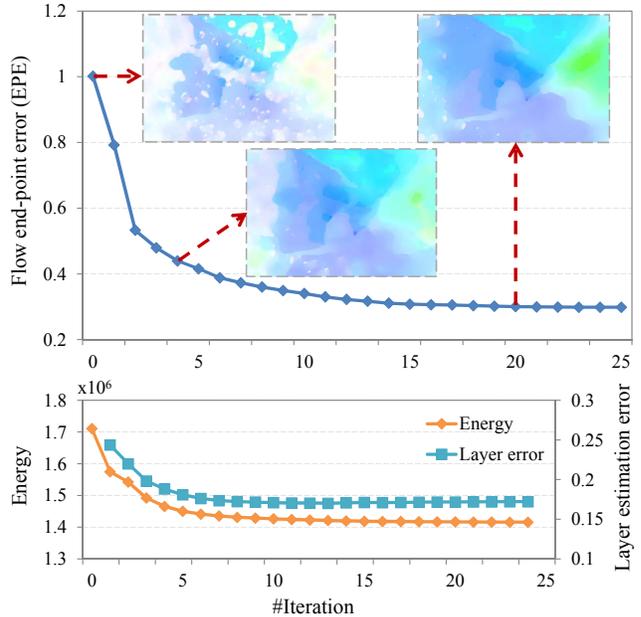


Figure 1: Convergence of the proposed method. Top: optical flow estimation error (EPE) *w.r.t.* iterations. Bottom: energy and layer estimation errors *w.r.t.* iterations. The layer error is evaluated as $1 - NCC(\text{GT } \mathbf{L}_2, \text{estimated } \mathbf{L}_2)$.

and real images (e.g. Middlebury [2] and Sintel [7] flow datasets, and the reflection dataset in [21]).

Initialization. Being an alternated method, the proposed algorithm requires an initialization to start the alternation. One can start from either an initial optical flow estimation or from an initial layer separation. The latter one is used in our experiments, and the initialization details will be given later in the experiments.

Parameters. In the following experiments, the weights of the priors, i.e. λ_L, λ_F , are roughly tuned according to the results. Both TV and TGV^2 flow regularizers worked well, consistently improving the accuracy upon initialization. Due to space limitation, in the following we report the results using TV (i.e. $k = 1$). The results using TGV^2 (i.e. $k = 2$) can be found in the *Supplementary Material*.

5.1. Static Foreground Cases

We start from the simpler case where only background layer \mathbf{L}_1 is dynamically changing by an unknown motion field \mathbf{U} , while the foreground layer is static (i.e. $\mathbf{L}_2 \equiv \mathbf{L}'_2$ and $\mathbf{V} \equiv \mathbf{0}$). The task is to estimate flow field \mathbf{U} and component layers $\mathbf{L}_1, \mathbf{L}'_1, \mathbf{L}_2$. Again, we would like to emphasize that, even though we call it the “simpler case”, to jointly estimate an accurate flow field and recover latent layers remains a challenging task. To the best of our knowledge,

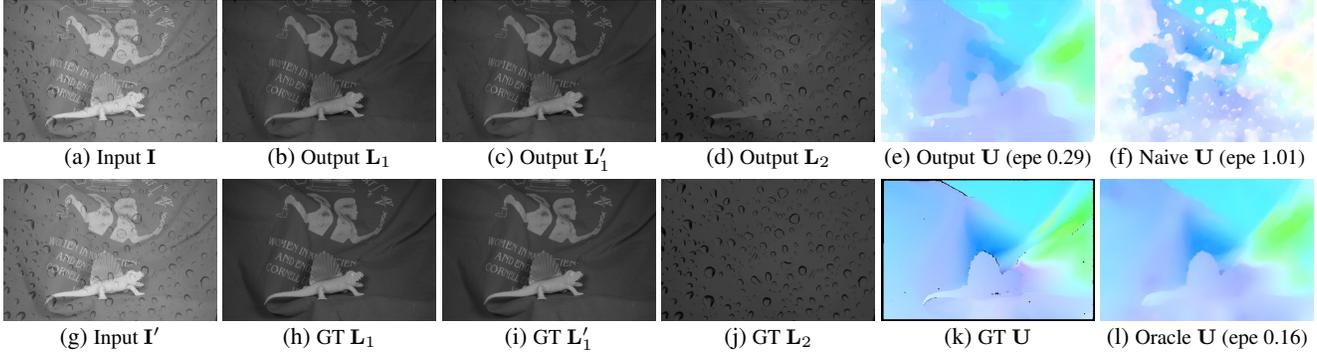


Figure 2: Performance evaluation of the proposed method on a single flow case, where a rain image is superimposed on the Dimetrodon image pair. The estimated flow (e) is significantly better than the initialization (f), a naive optical flow estimate without layer separation. The error evolution curve is shown in Fig. 1. Oracle flow (l) is computed with clean background images (*i.e.* with ground-truth layer separations). (**Best viewed on screen**)

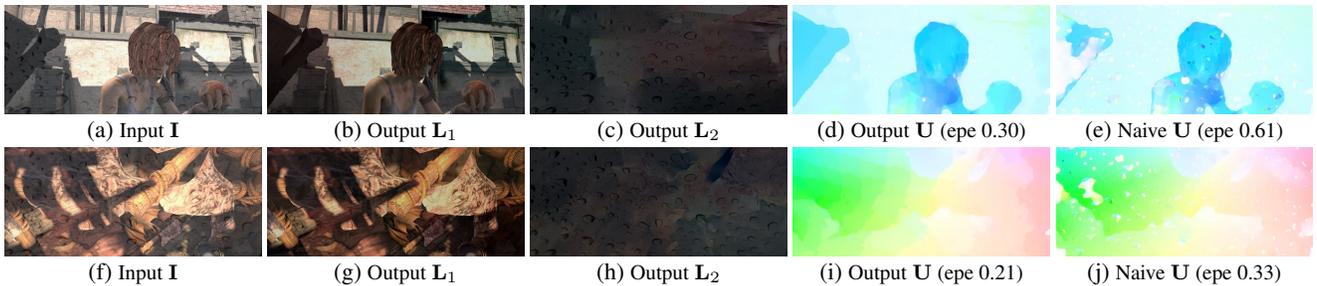


Figure 3: Typical results of our method on single-flow cases, where the rain drop image is superimposed on images from the Sintel dataset. For clarity, we only show here the first frame **I** and its layer separation result. (**Best viewed on screen**)

Table 1: Mean flow EPE for three Sintel image sequences superimposed with the static rain image. Oracle flows are computed with clean background images.

Sequence	Naive flow	Our flow	Oracle
“alley1”	0.49	0.35	0.22
“sleeping1”	0.80	0.33	0.12
“sleeping2”	0.26	0.21	0.07

there was no previous method that recovers both a complex dense flow field under transparency/reflection, and separate the two constituting layers.

In the following tests, a rather conservative strategy is used to initialize the proposed method: we initiate the static foreground image L_2 to be all zeros. Consequently, in the beginning of the optimization we compute an initial optical flow field naively based on the two input images.

Seeing through rain is a practical situation where measures should be taken to avoid the rain ruining vision systems. In the first test, we first synthesized a scene by superimposing a static rain image over the pair of Dimetrodon in the Middlebury dataset. Gray images were used. As illustrated in Fig. 1, within about 25 iterations, the optical flow

estimation error has been decreased from about 1.0 pixels to about 0.3 pixels. This demonstrates the advantage of our formulation for robust optical flow estimation. The qualitative results are demonstrated in Fig. 2.

Additionally, we overlay the rain image with three color image sequences from the Sintel dataset. We evenly sampled 10 images from the “alley 1”, “sleeping 1”, and “sleeping 2” sequences respectively, and Table 1 shows that the proposed method has clearly reduced the mean EPE of initial flows. Two typical results are shown in Fig. 3.

To further test the performance of our method, we synthesized another pair by superimposing the Lena image with the Grove image in the Middlebury dataset. The results are demonstrated in Fig. 4. Again, we obtained a much better optical flow compared to the initial naive optical flow estimate. As for the layer separation results, the portrait of Lena can be hardly seen in the restored grove images.

In Fig. 5 we show the image gradient statistics of the three foreground images used in the above experiments. The experimental results have shown that the proposed method works well on these images with the sparse gradient prior. Whenever available, other strong statistical priors can be incorporated into the optimization framework to further improve the performance.



Figure 4: Performance evaluation of the proposed method on a single flow case, where the Lena image is superimposed on the Grove image pair. The estimated flow (e) is significantly better than the initialization (f), a naive optical flow estimate without layer separation. Oracle flow (l) is computed with ground-truth L_2 . (*Best viewed on screen*)

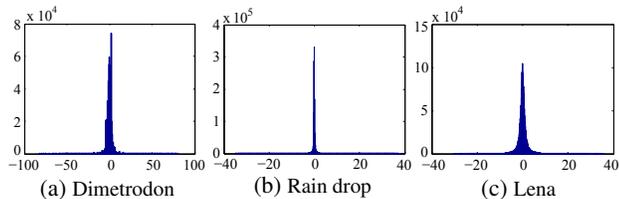


Figure 5: Gradient statistics of three used images.

5.2. Dynamic Foreground Cases

In this section, we test the proposed method in the dynamic foreground cases, where the task is that given two frames of input images I and I' , recover four component layers L_1, L'_1, L_2, L'_2 , and two dense motion fields U, V . In the problem of reflection removal, both the background scene and the reflection can be dynamic, which can give rise to such a situation.

We use two pairs of dynamic reflection scenes from [21] to test the proposed method on the double-layer optical flow problem. In previous single-flow experiments, we initialize the method with foreground layers being all zero. However, this simple strategy did not work for the double-flow case. No reasonably good flow field could be obtained with this strategy for the background or reflection layer, especially for the reflection layer as its signal is weak. Indeed, the fact that the background layer is much more prominent has been taken advantage of by some layer separation methods [21][14] which align the input images with respect to the background layer. To obtain proper initialization, we first ran method of [21] for initial layer separations², then computed initial optical flows on them.

The initial and final results are presented in Fig. 6. Vi-

²Method of [21] takes multiple images as input, with one of them being the reference on which the reflection is to be removed. We apply this method on two images, and run it twice with each image as reference.

Table 2: Mean image warping errors (in gray levels) from the double-flow estimation results.

Image pair	Initial results	Our final results
#1	6.27	2.55
#2	3.86	1.49

sually inspected, the final optical flow fields are smoother and more consistent (see e.g. the results on the back wall in the first example, and results on the floor in the second example). As no ground truth optical flow is available, we use image warping error to quantitatively evaluate the estimated flows. The warping error for a pixel \mathbf{x} in L_1 or L_2 is $\|L_1(\mathbf{x} + U(\mathbf{x})) - L'_1(\mathbf{x})\|_2$ or $\|L_2(\mathbf{x} + V(\mathbf{x})) - L'_2(\mathbf{x})\|_2$, respectively. We compute the mean warping errors for all pixels on L_1 and L_2 . As shown in Table 2, our method has significantly reduced the warping error upon the initializations. Figure 6 shows the improvements of the reflection removal results upon the initial estimates.

Discussion. The dynamic foreground case with double-layer flow estimation is generally much harder than the single-flow case. This is not only because the former has more unknown variables to be solved for, but also due to the difficulties in obtaining a good initialization. Nevertheless, our experiments show that the proposed method consistently improved the reasonable initializations given to it, for both the single-flow and double-flow cases.

Limitation. The proposed method is better suited for scenarios where the correlation between latent layers and their flow fields are relatively small. It will fail if both the two layers are textureless (as infinite numbers of possible motions exist satisfying the BCC constraints), or they undergo a same motion (thus the original BCC holds and only a single motion field can be extracted).

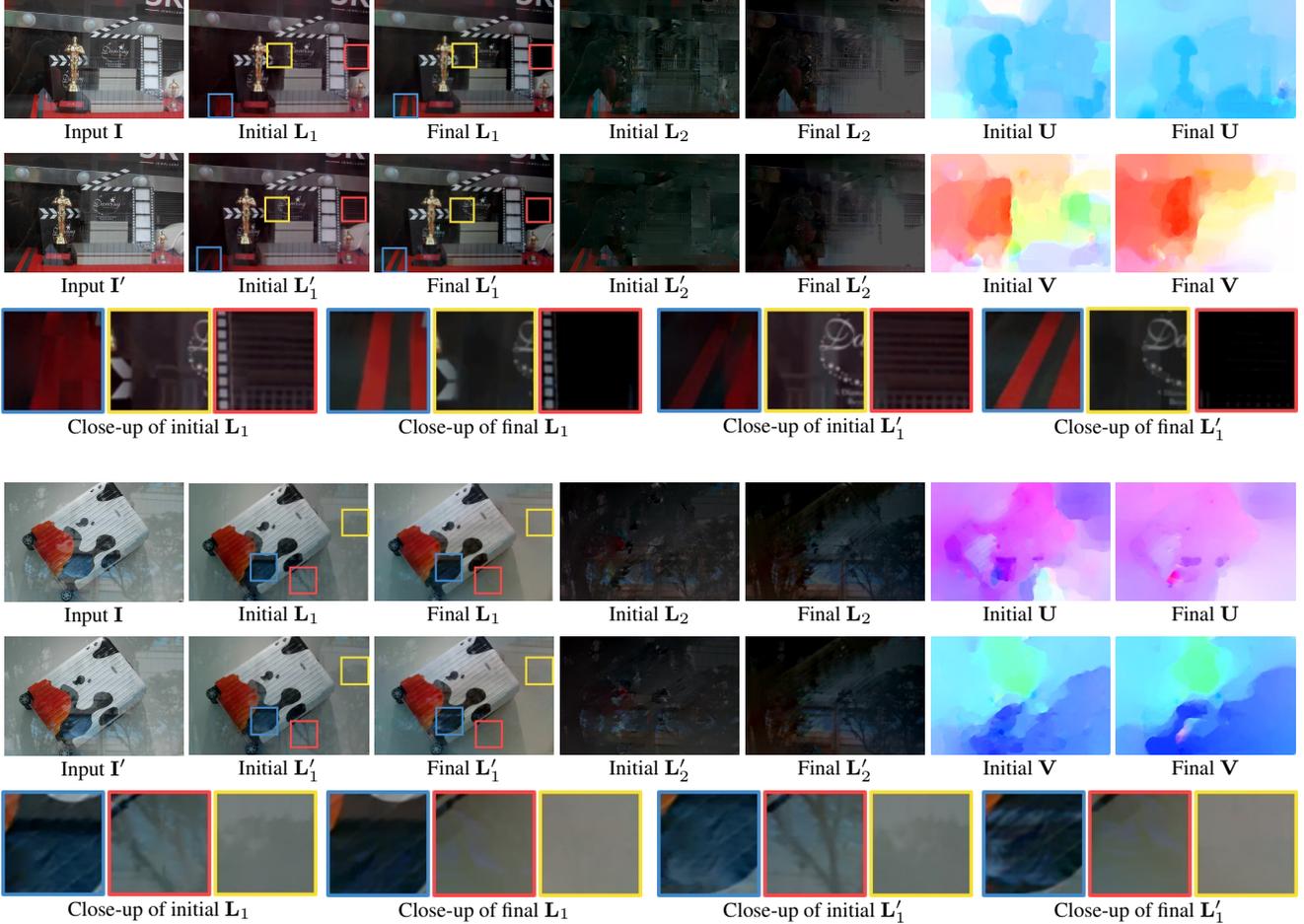


Figure 6: Experimental results on real reflection images. The initial layer separations are estimated by running method of [21] on the two input images. Visually inspected, the final optical flow fields are smoother and more consistent (see *e.g.* the results on the back wall in the first example, and results on the floor in the second example). The corresponding warping errors are presented in Table 2. The close-up images in the third rows show the improvements of the reflection removal results upon the initial estimates. (*Best viewed on screen*)

6. Conclusions and Future Work

This paper has defined the problem of robust optical flow estimation in the presence of possibly moving transparent or reflective layers. To our knowledge, the problem goes beyond the scope of conventional optical flow methods and was not properly investigated before.

We have presented a generalized double-layer brightness constancy condition as well as an optimization framework to solve this problem. The double-layer brightness constancy condition couples the flow fields and the brightness layers. Encouraging experimental results of optical flow estimation and layer separation on challenging data have been obtained, even though we are using simple priors for them. We hope that this paper can inspire future works to further address this challenging ill-posed problem.

Our current framework is based on a generative model,

which is applied uniformly to both the foreground and background layers. In future, we plan to leverage discriminative models to exploit the differences between the two layers for better layer separation. We also would like to explore some other optical flow priors. One possible strategy is to apply piecewise parametric motion model [16, 44], which provides stronger constraints than general smoothness regularizers such as a TV, and is recently demonstrated to have advanced performances [44]. Some other issues such as occlusion handling could also be considered.

Acknowledgments H. Li’s research is funded in part by Australian Research Council Grants of DP120103896, LP100100588, ARC Centre of Excellence on Robotic Vision (CE140100016) and NICTA (Data61). Y. Dai’s contribution is funded in part by ARC Grants (DE140100180, LP100100588) and National Natural Science Foundation of China (61420106007).

References

- [1] V. Auvray, P. Bouthemy, and J. Liénard. Joint motion estimation and layer segmentation in transparent image sequences: application to noise reduction in X-ray image sequences. *EURASIP Journal on Advances in Signal Processing*, pages 19:1–19:21, 2009. 2
- [2] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision (IJCV)*, 92(1):1–31, 2011. 5
- [3] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, (9):886–896, 1992. 2, 4
- [4] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding (CVIU)*, (1):75–104, 1996. 2
- [5] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010. 4
- [6] T. Brox, A. Bruhn, N. Papenbergh, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, pages 25–36, 2004. 3
- [7] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conference on Computer Vision (ECCV)*, pages 611–625, 2012. 5
- [8] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision (JMIV)*, 40(1):120–145, 2011. 5
- [9] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3869–3872, 2008. 5
- [10] Y. Chen, T. Chang, C. Zhou, and T. Fang. Gradient domain layer separation under independent motion. In *International Conference on Computer Vision (ICCV)*, pages 694–701, 2009. 2
- [11] T. Darrell and E. Simonecelli. ‘Nulling’ filters and the separation of transparent motions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 738–739, 1993. 2
- [12] H. Farid and E. H. Adelson. Separating reflections and lighting using independent components analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 262–267, 1999. 2
- [13] K. Gai, Z. Shi, and C. Zhang. Blind separation of superimposed moving images using image statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(1):19–32, 2012. 2
- [14] X. Guo, X. Cao, and Y. Ma. Robust separation of reflection from multiple images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2195–2202, 2014. 2, 7
- [15] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision (IJCV)*, 12(1):5–16, 1994. 2
- [16] S. X. Ju, M. J. Black, and A. D. Jepson. Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 307–314, 1996. 2, 8
- [17] K. Langley, D. Fleet, and T. Atherton. Multiple motions from instantaneous frequency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 846–849, 1992. 2
- [18] K. Langley, D. J. Fleet, and T. J. Atherton. On transparent motion computation. In *British Machine Vision Conference (BMVC)*, pages 247–256, 1992. 2
- [19] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(9):1647–1654, 2007. 2, 3
- [20] A. Levin, A. Zomet, and Y. Weiss. Learning to perceive transparency from the statistics of natural scenes. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1247–1254, 2002. 2
- [21] Y. Li and M. Brown. Exploiting reflection change for automatic reflection removal. In *International Conference on Computer Vision (ICCV)*, pages 2432–2439, 2013. 2, 5, 7, 8
- [22] Y. Li and M. S. Brown. Single image layer separation using relative smoothness. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2752–2759, 2014. 2, 3
- [23] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Z. Zhou, and L.-F. Cheong. Simultaneous video defogging and stereo reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4988–4997, 2015. 2
- [24] M. Pingault and D. Pellerin. Optical flow constraint equation extended to transparency. In *European Signal Processing Conference*, pages 1–4, 2002. 2
- [25] A. Ramirez-Manzanares, M. Rivera, P. Kornprobst, and F. Lauze. Multi-valued motion fields estimation for transparent sequences with a variational approach. *INRIA Technical Report*, 2006. 2
- [26] B. Sarel and M. Irani. Separating transparent layers through layer information exchange. In *European Conference on Computer Vision (ECCV)*. 2
- [27] Y. Y. Schechner, N. Kiryati, and R. Basri. Separation of transparent layers using focus. *International Journal of Computer Vision (IJCV)*, 39(1):25–39, 2000. 2
- [28] M. Shizawa and K. Mase. Simultaneous multiple optical flow estimation. In *International Conference on Pattern Recognition (ICPR)*, pages 274–278, 1990. 2, 4
- [29] M. Shizawa and K. Mase. Unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 289–295, 1991. 2
- [30] F. Steinbrücker, T. Pock, and D. Cremers. Large displacement optical flow computation without warping. In *International Conference on Computer Vision (ICCV)*, pages 1609–1614, 2009. 5

- [31] D. Sun, S. Roth, and M. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision (IJCV)*, 106(2):115–137, 2014. 2
- [32] D. Sun, E. B. Sudderth, and M. J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2226–2234. 2
- [33] R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 246–253, 2000. 2
- [34] R. Szeliski and P. Golland. Stereo matching with transparency and matting. In *International Conference on Computer Vision (ICCV)*, pages 517–524, 1998. 2, 3
- [35] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(9):1459–1472, 2005. 2, 3
- [36] J. Toro, F. J. Owens, and R. Medina. Multiple motion estimation and segmentation in transparency. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 6, pages 2087–2090, 2000. 2
- [37] Y. Tsin, S. B. Kang, and R. Szeliski. Stereo matching with linear superposition of layers. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(2):290–301, 2006. 2
- [38] J. Y. Wang and E. H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing (TIP)*, 3(5):625–638, 1994. 2
- [39] Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 520–526, 1997. 2
- [40] Y. Weiss. Deriving intrinsic images from image sequences. In *International Conference on Computer Vision (ICCV)*, pages 68–75, 2001. 2
- [41] Y. Wexler, A. Fitzgibbon, and A. Zisserman. Bayesian estimation of layers from multiple images. In *European Conference on Computer Vision (ECCV)*, pages 487–501, 2002. 2
- [42] J. Wulff and M. J. Black. Modeling blurred video with layers. In *European Conference on Computer Vision (ECCV)*, pages 236–252, 2014. 2
- [43] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman. A computational approach for obstruction-free photography. *ACM Transactions on Graphics (TOG)*, 34(4):79, 2015. 2
- [44] J. Yang and H. Li. Dense, accurate optical flow estimation with piecewise parametric model. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1019–1027, 2015. 2, 8
- [45] G. Ye, E. Garces, Y. Liu, Q. Dai, and D. Gutierrez. Intrinsic video and applications. *ACM Transactions on Graphics (TOG)*, 33(4):80, 2014. 2
- [46] S.-K. Yeung, T.-P. Wu, and C.-K. Tang. Extracting smooth and transparent layers from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2008. 2, 3
- [47] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Annual Symposium of the German Association for Pattern Recognition (DAGM)*, pages 214–223, 2007. 3, 4, 5