Computer Vision and Image Understanding 116 (2012) 250–261

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Improving sub-pixel correspondence through upsampling

Li Xu [a], Jiaya Jia [a,*], Sing Bing Kang [b]

[a] The Chinese University of Hong Kong, Hong Kong
[b] Microsoft Research, Redmond, WA, USA

ARTICLE INFO

Article history:
Received 31 October 2009
Accepted 10 November 2011
Available online 20 November 2011

Keywords:
Sub-pixel correspondence
Stereo matching
Optical flow
Interpolation error
Regularization bias

ABSTRACT

Many fundamental computer vision problems, including optical flow estimation and stereo matching, involve the key step of computing dense color matching among pixels. In this paper, we show that by merely upsampling, we can improve sub-pixel correspondence estimation. In addition, we identify the *regularization bias* problem and explore its relationship to image resolution. We propose a general upsampling framework to compute sub-pixel color matching for different computer vision problems. Various experiments were performed on motion estimation and stereo matching data. We are able to reduce errors by up to 30%, which would otherwise be very difficult to achieve through other conventional optimization methods.

1. Introduction

Computing a dense displacement map between two images is a fundamentally important step in solving many computer vision problems, such as stereo matching and optical flow estimation. Techniques for improving correspondences are usually based on model or prior refinement, such as introducing outlier rejection [24,4], 2nd order regularization [28,27], and edge preserving smoothness [29], or on improving the optimization method itself [16,24,28,17].

The coarse-to-fine strategy for motion estimation is popular [5,6]; apart from helping avoid local minima, it also helps reduce the search space. However, this strategy is mostly treated as a numerical approximation. In this paper, we suggest that an unconventional *fine-to-coarse* strategy is very useful in establishing more accurate sub-pixel correspondence. We show, through theoretical analysis and empirical validation, that this intuition is correct. This paper is also the first attempt to fuse results in different scales for effective error suppression.

We analyze the effect of image scale in sub-pixel correspondence estimation, and propose a general and unified upsampling framework to improve it. This framework is applicable to different tasks such as motion estimation and stereo matching. It has three main contributions. First, we analyze from a quantization error perspective how changing image scale would affect sub-pixel correspondence. Second, we study the energy bias caused by the regularization term which was generally neglected in previous work. Third, we propose a unified upsampling framework that fuses the matching results in different scales to further suppress errors, based on the observation that upsampling improves sub-pixel correspondence within smooth regions but not necessarily at region boundaries with abrupt changes in color. Our framework is simple and general, and can be employed in stereo matching and flow estimation to improve result quality.

In our experiments, consistent matching accuracy enhancement is observed where up to 30% error reduction is yielded. Given the simplicity and the generality of our method, it is noteworthy that this substantial melioration would otherwise be very difficult to be obtained even by specifically adjusting previous matching schemes.

Our system is different from optical flow based super-resolution [1,32,10], which focuses on recovering high-resolution images making use of optical flow. These methods do not discuss whether high-resolution images help improve the flow estimate or not. Methods that use interpolation to enhance the accuracy in computing data costs [3,23,25,18,19] also do not discuss the effect of the regularization terms and still solve the final energy functions at the original resolution. In contrast, we show that performing matching with high-res images not only reduces the data cost, but as well decreases inherent regularization bias, resulting in an explicit upsampling framework to find sub-pixel correspondence.

2. Related work

We briefly review approaches of stereo matching and optical flow estimation in terms of ways to improve the optimization

* Corresponding author.
E-mail address: leojia@cse.cuhk.edu.hk (J. Jia).

1077-3142/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2011.11.003

methods (the solver) or to refine the matching energy cost and smoothness prior (the model).

Two-view stereo matching is generally a 1D matching problem using disparities [21]. Discrete optimization, such as belief propagation [24], graph-cuts [16,13], and the QPBO algorithm [28], was proposed to solve the problem.

Several methods improve the sub-pixel accuracy for discrete disparity estimation. The sub-pixel displacement is usually estimated by parabola fitting of neighboring integer values [26]. A classification of early sub-pixel estimation methods using peak detection and interpolation was presented in [9]. Shimizu and Okutomi [23] computed the displacement for several half-pixel-shifted image pairs. Birchfield and Tomasi [3] improved the discrete data term by sampling and compared the reference image with a linearly interpolated matching view. In [25], Szeliski and Scharstein proposed using signal interpolation, interval matching, and quadratic fitting to infer the continuous pixel matching cost in the disparity space. To improve the sub-pixel accuracy of stereo matching, in [18,19], spline interpolation followed by symmetric refinement and improved correlation-based data cost measure were respectively adopted. Gehrig and Franke [11] proposed explicit evaluation of the data cost at fractional disparities and applied disparity smoothing between adjacent pixels.

Though interpolation techniques were used to improve data cost computation, above methods solve the matching problem still at the original resolution. How regularization term affects the sub-pixel accuracy has generally been ignored. In comparison, we contribute an explicit upsampling framework with a refinement fusion step. Our framework not only reduces data matching cost, but as well decreases the bias caused by enforcing regularization.

Optical flow estimation aims to compute 2D apparent motion for each pixel. Variational methods are dominant in this field [14,4,6]. They have been extended by incorporating different priors such as color segmentation [33,30]. Coarse-to-fine strategies were typically adopted to avoid local minima and for speedup [5,6]. An extensive evaluation was presented in [2]. An important issue is to improve sub-pixel accuracy by refining the data cost through interpolation [5,6,30].

With multi-scale pixel correspondences, optical flow based super-resolution [1,32,10] links motion estimation to resolution enhancement to recover high-res images. Iterations between super-resolution and flow estimation are performed. As high-res images are only used to compute flow, how resolution change affects the flow results is still unknown.

In this paper, we do not attempt to use multiple low-resolution images to construct a single high-resolution one. Instead, we propose upscaling input images and discuss how this scheme would affect the matching accuracy.

3. Sub-pixel correspondence

Our goal is sub-pixel correspondence across two views. The dense displacement field (or disparity map) $\mathbf{d}^*$ between the reference view I_0 and the target view I_1 is constructed by solving an optimization problem expressed as

$$\mathbf{d}^* = \text{argmin}_{\mathbf{d}} E_{data}(\mathbf{d}) + \lambda E_{reg}(\mathbf{d}), \tag{1}$$

where $E_{data}(\mathbf{d})$ is a data term and $E_{reg}(\mathbf{d})$ is a regularization term. $E_{data}(\mathbf{d})$ is generally defined as minimizing color difference between corresponding pixels/sub-pixels, written as

$$E_{data}(\mathbf{d}) = \sum_{\mathbf{m}} \rho(\|I_1(\mathbf{m} + \mathbf{d}_I(\mathbf{m})) - I_0(\mathbf{m})\|^2), \tag{2}$$

where $\rho(\cdot)$ is a monotonically increasing robust function to reject outliers. $\mathbf{d}_I(\mathbf{m}) : \mathbb{Z}^2 \mapsto \mathbb{R}^2$ is the dense displacement (or motion) field, with the same resolution as the image. There are other ways to define the color similarity, such as varying the robust function [21,4] or locally grouping the color difference [8,31]. However, most of these methods are variations of Eq. (2). Accurate disparity or flow field estimation depends largely on the data costs.

4. Sub-pixel correspondence accuracy

We start by raising the question of whether data term computation on a ground truth high-resolution image pair is more accurate than that on low-resolution images or not. Since ideal high-resolution images are the goal of many image super-resolution and upsampling approaches, we call them *ideal upsampled* images. We review the image formation process in order to establish the relationship between the low-resolution inputs and their ideal upsampled versions. Note that practically, there may not exist ideal upsampled images if only one resolution is given. Our experiments show that in this case, advanced upsampling methods can be adopted to similarly improve sub-pixel matching accuracy, as will be detailed in Sections 6 and 7.

4.1. Image formation process

We denote by $L(\mathbf{m})$ the input coarse-scale image, where $\mathbf{m} \in \mathbb{Z}^2$ is the 2D coordinate. It can be modeled as an integration over a light field, i.e.,

$$L(\mathbf{m}) = \int_{\mathbf{x} \in S(\omega)} E(\mathbf{x}) \omega(\mathbf{x} - \mathbf{m}) d\mathbf{x}, \tag{3}$$

where $E(\mathbf{x})$ is a continuous function representing the radiance at any location $\mathbf{x}$.ω is the point spread function (PSF). $S(\omega)$ is the support region of the PSF.

By discretizing the continuous function $E(\cdot)$ into image $H(\mathbf{p})$, we obtain

$$L(\mathbf{m}) = \sum_{\mathbf{p} \in S(W, \mathbf{m})} W(\mathbf{m}, \mathbf{p}) \cdot H(\mathbf{p}), \tag{4}$$

where $W(\mathbf{m}, \mathbf{p})$ is a discrete weight function dependent of the continuous PSF $\omega(\cdot)$. By the energy preservation rule, $\sum_p W(\mathbf{m}, \mathbf{p}) = 1$. Eq. (4) relates an ideal upsampled image $H(\mathbf{p})$ to its coarse-scale version $L(\mathbf{m})$.$\mathbf{p}$ and $\mathbf{m}$ are coordinates in L and H, respectively. $S(W, \mathbf{m})$ is the support region of W at point $\mathbf{m}$. The model in (4) can be abstracted as prefiltering followed by decimation [12]:

$$L = \mathcal{D} \cdot \mathcal{B} \cdot H, \tag{5}$$

where $\mathcal{B}$ and $\mathcal{D}$ are respectively the prefilter (to prevent aliasing) and decimation processes. For simplicity, we denote $BH = \mathcal{B} \cdot H$. Decimation of $\mathcal{D}$ samples pixels to generate a coarser level image. We define $\mathbf{p} = \mathbf{m} \cdot s$, where s is the scale factor.

4.2. Accuracy analysis with decimation

The only difference between L and BH is pixel decimation, that is $L(\mathbf{m}) = BH(\mathbf{m} \cdot s)$. We analyze the matching process for these two images with their data terms similarly defined as:

$$\sum_{\mathbf{p}} \rho(\|BH_1(\mathbf{p} + \mathbf{d}_{BH}(\mathbf{p})) - BH_0(\mathbf{p})\|^2),$$
$$\sum_{\mathbf{m}} \rho(\|L_1(\mathbf{m} + \mathbf{d}_L(\mathbf{m})) - L_0(\mathbf{m})\|^2),$$

where BH_0 and BH_1 are the prefiltered reference and target images respectively. Given the scale difference between BH and L, we have

$$\mathbf{d}_L(\mathbf{m}) = \frac{1}{s} \mathbf{d}_{BH}(\mathbf{m} \cdot s), \tag{6}$$

where $\mathbf{p} = \mathbf{m} \cdot s$. This difference causes discrepancy in computing the data term.

To facilitate discussion, we define a latent continuous light field $I(\mathbf{x})(\mathbf{x} \in \mathbb{R}^2)$ from which BH and L are sampled. The sampling intervals of BH and L are h and $s \cdot h$, respectively. So, we have

$$\begin{aligned} \mathrm{BH}(\mathbf{p}) &= I(\mathbf{p} \cdot h), \\ \mathrm{L}(\mathbf{m}) &= I(\mathbf{m} \cdot s \cdot h), \end{aligned} \tag{7}$$

where $\mathbf{p}$ and $\mathbf{m}$ are integers.

It is notable that $I(\mathbf{x})$ is unknown. It is shown here only for helping analyze how using BH and L introduce errors in data cost estimation especially for points $\mathbf{x}$ that are in between the samples. These points have fractional coordinates and the commonest strategy to infer their intensities or colors is by interpolation of neighboring values. Our analysis will focus on how the interpolation errors, defined as $e_{\mathrm{L}}(\mathbf{x}) = \mathrm{L}(\mathbf{x}) - I(\mathbf{x})$ and $e_{\mathrm{BH}}(\mathbf{x}) = \mathrm{BH}(\mathbf{x}) - I(\mathbf{x})$ for L and BH respectively, vary. It is obvious that e equals 0 for all points with integer coordinates and is non-zero otherwise. To make the theoretical discussion specific, we consider bilinear interpolation for obtaining sub-pixel colors.

Claim 1. *For an underlying differentiable light field, given small sampling interval h and linear interpolation, we have the average interpolation error* $(e_{BH}) \leqslant (e_L)$.

Proof. We prove it in 1D, as the 2D derivation is similar. For any sub-pixel at location x, where x is not an integer, its color is interpolated from nearby pixels $p_0 = \lfloor \frac{x}{h} \rfloor h$ and $p_1 = \lceil \frac{x}{h} \rceil h$. We define $\Delta x = x - p_0$, where $\Delta x \in [0, h)$. Using linear interpolation, the approximated color for BH at location x is given by

$$\mathrm{BH}(x) = \frac{h - \Delta x}{h} I(p_0) + \frac{\Delta x}{h} I(p_1). \tag{8}$$

By substituting Eq. (8) into e_{BH} and taking the Taylor expansion at x, we obtain

$$\begin{aligned} e_{\mathrm{BH}} &= \mathrm{BH}(x) - I(x) = \frac{h - \Delta x}{h} I(p_0) + \frac{\Delta x}{h} I(p_1) - I(x) \\ &= \frac{1}{2} \ddot{I}(x) \Delta x (h - \Delta x) + \frac{1}{3!} \dddot{I}(x) \Delta x (h - \Delta x)(h - 2\Delta x) + \mathrm{O}(h^4), \end{aligned} \tag{9}$$

where $\ddot{I}(x)$ and $\dddot{I}(x)$ are the second-order and third-order derivatives of the light field at x in BH.

Similarly, for L, we can express error e_{L} as

$$\begin{aligned} e_{\mathrm{L}} = \frac{1}{2} \ddot{I}(x) \Delta x' (sh - \Delta x') + \frac{1}{3!} \dddot{I}(x) \Delta x (sh - \Delta x')(h - 2\Delta x') \\ + \mathrm{O}(h^4), \end{aligned} \tag{10}$$

where $\Delta x' \in [0, sh)$. Eqs. (9) and (10) indicate that the interpolation error approaches zero when the second- and higher-order derivatives of the irradiance are very small locally. It would happen with very low frequency content, which makes the two interpolation errors equal zeros.

For non-zero interpolation errors, the ratio of the average errors can be expressed as

$$r = \sqrt{\frac{\frac{1}{h}\int_0^h (e_{\mathrm{BH}}(\Delta x))^2 d\Delta x}{\frac{1}{sh}\int_0^{sh} (e_{\mathrm{L}}(\Delta x'))^2 d\Delta x'}} = \sqrt{\frac{c_1 h^4 + c_2 h^6 + \mathrm{O}(h^7)}{c_1 (sh)^4 + c_2 (sh)^6 + \mathrm{O}(h^7)}}, \tag{11}$$

where $c_1 = \frac{\ddot{I}(x)}{2!30}$ and $c_2 = \frac{\dddot{I}(x)}{3!210}$. Note that

$$\lim_{h \to 0} r = \frac{1}{s^2}. \tag{12}$$

This proves that lower resolution images yield larger interpolation errors on average. The derivation in 2D is similar, with the corresponding ratio being $1/s^4$ instead. □

The above analysis shows that higher-resolution images can make data cost estimation more accurate compared to their coarser-level versions. When the sampling rate is high enough with respect to the local intensity variation, the error bound in Eq. (12) holds.

The ratio r for other interpolation methods can also be derived. We omit the proof here as they are quite similar to the one shown above. The ratio r using the nearest neighbor (NN) interpolation is proportional to $1/s^2$ as the interpolation error for NN is $\mathrm{O}(h)$. Polynomial interpolation with degree n has the interpolation error $\mathrm{O}(h^n)$. We conclude that if a polynomial interpolation method is used for sub-pixel matching, performing the same procedure on a higher resolution image would generally produce a more accurate color difference measure.

There are special cases for Eq. (11). (a) If $c_1 = 0$ and the high order terms are not neglectable (e.g. the inflection points of local irradiance), the ratio is dominated by the high-order terms. Thus the ratio r could be even smaller. (b) If h is significant with respect to the spatial color variation, or the underlying light field is not differentiable at a step edge, there is no theoretical bound for r. We will compare the interpolation errors empirically.

We experimented with 24 images from the Kodak color image set (http://www.r0k.us/graphics/kodak/). All images in their original resolution approximate light field I. They are respectively downsampled by filtering and are decimated to generate two images BH and L with scale difference $s = 2$. The pixel intensity is scaled to $[0, 1]$. One example is shown in Fig. 1.

We first interpolate intensities at the sub-pixel locations of L, which are also the sub-pixel locations of BH. The average interpolation errors (AIEs) $|e_{\mathrm{BH}}|$ and $|e_{\mathrm{L}}|$ are calculated on these common locations, listed in Table 1. AIEs of BH are consistently smaller than those yielded on L for the three interpolation methods we experimented with.

The proof of Claim 1 also indicates that interpolation errors produced from higher-resolution images are generally smaller for smoothly varying signals, but not for signals with abrupt change necessarily. To illustrate it, we select two patches (size 100×100) highlighted in yellow in Fig. 1 and compare their mean interpolation errors counting in the influence of local structure and texture. We sum the squared interpolation errors ${e_{\mathrm{L}}}^2$ and ${e_{\mathrm{BH}}}^2$ for each patch and compute ratio $r' = \sqrt{\sum e_{\mathrm{BH}}^2 / \sum e_{\mathrm{L}}^2}$.

The upper patch contains high-frequency textures, resulting in ratio $r' = 0.49$. Interpolation errors near step edges are relatively large. The lower patch contains a larger degree of low-frequency color transition, which produces a smaller ratio $r' = 0.26$. These values indicate *if spatial color variation is small, using prefiltered high-resolution images effectively reduce interpolation errors*. We use this principle to guide displacement refinement in Section 6.

4.3. Accuracy analysis with prefiltering

We have shown that in computing matching cost, the prefiltered image BH generally produces smaller interpolation errors than L. In this section, we demonstrate that matching on H is even better in terms of robustness to find pixel correspondences.

It is explainable that pixel matching on H outperforms that on BH. On the one hand, a blurred image has less high frequency details, making matching more ambiguous. On the other hand, as we will show shortly, when the local displacement field is flat, pixel correspondence estimation between image pair H_0 and H_1 is more stable.

We denote the pixel color difference on the sharp and blurred image pair as

$$\mathcal{M}(\mathbf{d}_{\mathrm{H}}, \mathbf{p}) = \mathrm{H}_1(\mathbf{p} + \mathbf{d}_{\mathrm{H}}) - \mathrm{H}_0(\mathbf{p}), \tag{13}$$

and

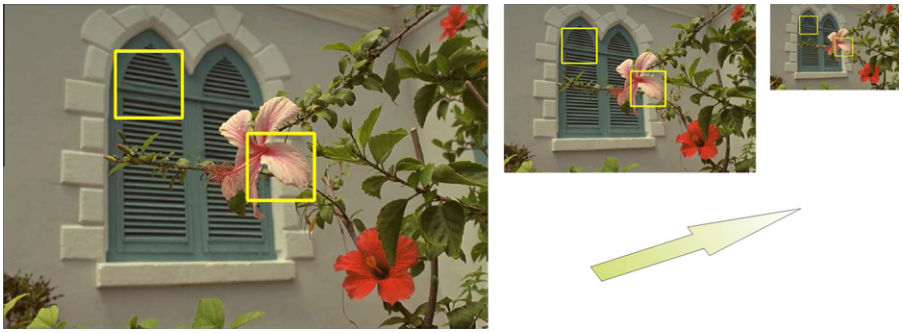

Fig. 1. One example from the Kodak color image set. The high-resolution image (left) simulates I. It is downsampled to produce two images (right) in different scales. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Average errors using three interpolation methods.

Method	Mean $\|e_L\|$	Mean $\|e_{BH}\|$
Nearest neighbour	0.0909	0.0468
Bilinear	0.0769	0.0370
Bicubic	0.0803	0.0383

$$\mathcal{BM}(\mathbf{d}_{\mathrm{BH}},\mathbf{p}) = \mathrm{BH}_1(\mathbf{p}+\mathbf{d}_{\mathrm{BH}}) - \mathrm{BH}_0(\mathbf{p}),$$

respectively. If the displacement field varies slowly in a local region (which is very common), we can approximate

$$\mathcal{BM}(\mathbf{d}_{\mathrm{BH}},\mathbf{p}) \approx \sum_{\mathbf{p}'\in S(W,\mathbf{m})} \delta_{\mathbf{p}'}(\mathrm{H}_1(\mathbf{p}'+\mathbf{d}_{\mathrm{H}}) - \mathrm{H}_0(\mathbf{p}')), \tag{14}$$

where $\delta_{\mathbf{p}'}$ is the filter for pixel $\mathbf{p}'$, i.e., $W(\mathbf{x},\mathbf{p}') = \delta_{\mathbf{p}'}$. Considering a small displacement increment $\Delta\mathbf{d} > \mathbf{0}$ near the optimal point $\mathbf{d}^*$ for each pixel. If $\mathcal{M}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}) \approx \mathcal{M}(\mathbf{d}^*,\mathbf{p})$, the optimum can be easily mistaken as $\mathbf{d}^* + \Delta\mathbf{d}$ with a small perturbation of the solution. We compare solution stability in this regard for the original and filtered images in order to analyze how easy the results are influenced. The following inequality holds:

$$\begin{aligned}&\sum_{\mathbf{p}} \|\mathcal{M}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}) - \mathcal{M}(\mathbf{d}^*,\mathbf{p})\|^2 \\ &\geqslant \sum_{\mathbf{p}} \| \sum_{\mathbf{p}'\in S(W,\mathbf{m})} \delta_{\mathbf{p}'}(\mathcal{M}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}') - \mathcal{M}(\mathbf{d}^*,\mathbf{p}))\|^2,\end{aligned} \tag{15}$$

based on the Cauchy-Schwarz inequality. The right hand side of (15) can be further written as $\sum_{\mathbf{p}}\|\mathcal{BM}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}) - \mathcal{BM}(\mathbf{d}^*)\|^2$, yielding

$$\begin{aligned}&\sum_{\mathbf{p}} \|\mathcal{M}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}) - \mathcal{M}(\mathbf{d}^*,\mathbf{p})\|^2 \\ &\quad\geqslant \sum_{\mathbf{p}} \|\mathcal{BM}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}) - \mathcal{BM}(\mathbf{d}^*,\mathbf{p})\|^2.\end{aligned} \tag{16}$$

Eq. (16) shows if pixel matching on the sharp images is easily perturbed near the optimal point, i.e., $\mathcal{M}(\mathbf{d}^*+\Delta\mathbf{d},\mathbf{p}) \approx \mathcal{M}(\mathbf{d}^*,\mathbf{p})$, so does matching on the corresponding filtered images. It also implies that computing color difference on the ideal upsampled images can be more, or at least equally, robust than that on the filtered images when the displacement field is locally smooth.

5. Regularization bias

We also analyze the influence of image scales in computing the regularization cost. The regularization term in general enforces smoothness when the data costs for different labels are not distinctive enough especially in the textureless or occluded regions. The side effect includes the possible bias towards unfaithful or excessive smoothness account for using only local information. In what follows, we first quantitatively analyze the bias caused by regularization, and then discuss how image scales affect it.

One common choice of the regularization term is

$$\mathrm{E}_{reg}(\mathbf{d}) = \sum_{\mathbf{m}} \rho(\|\nabla\mathbf{d}(\mathbf{m})\|^2). \tag{17}$$

The gradient operator $\nabla\mathbf{d}(\mathbf{m})$ can be discretized with forward difference. To enable quantitative analysis, we use the example of the total variation regularizer [6], which simplifies Eq. (17) to

$$\mathrm{E}_{reg}(\mathbf{d}) = \sum_{(\mathbf{m},\mathbf{n})\in\mathcal{N}} \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{n})\|, \tag{18}$$

where $\mathcal{N}$ contains all neighboring pixel pairs. We define the regularization bias as the difference between the ground-truth displacement $\tilde{\mathbf{d}}$ and the displacement $\mathbf{d}_{reg}$ computed using E_{reg} for point $\mathbf{m}$. It is expressed as

$$bias(\mathbf{m}) = \|\tilde{\mathbf{d}}(\mathbf{m}) - \mathbf{d}_{reg}(\mathbf{m}|\{n\})\|. \tag{19}$$

$\mathbf{d}_{reg}(\mathbf{m}|\{\mathbf{n}\})$ is the smoothness cost for point $\mathbf{m}$, which depends on the disparity of the neighboring points, written as

$$\mathbf{d}_{reg}(\mathbf{m}|\{\mathbf{n}\}) = \operatorname{argmin} \sum_{(\mathbf{m},\mathbf{n})\in\mathcal{N}} \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{n})\|, \tag{20}$$

where $\mathbf{n}$ indexes the neighboring pixels of $\mathbf{m}$.

Fig. 2 shows an 1D illustration of the bias. There are three consecutive points in (a), based on which we analyze the regularization bias of the central point $\mathbf{m}$ with regard to $\mathbf{a}$ and $\mathbf{b}$.$\mathbf{d}(\mathbf{a})$ and $\mathbf{d}(\mathbf{b})$ have values 0.5 and 0.3 respectively in this example. Fig. 2c shows the regularization energy. The blue and red curves show how the energy changes with respect to different $\mathbf{d}(\mathbf{m})$ using the total variation regularizers $\|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a})\|$ and $\|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b})\|$ for $\mathbf{a}$ and $\mathbf{b}$ respectively. The dashed curve represents the regularization term $\mathbf{d}_{reg}(\mathbf{m})$, written as

$$\mathbf{d}_{reg}(\mathbf{m}|\{\mathbf{a},\mathbf{b}\}) = \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a})\| + \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b})\|. \tag{21}$$

It is clear that the regularization energy is minimized when $\mathbf{d}(\mathbf{m}) \in [\mathbf{d}(\mathbf{a}),\mathbf{d}(\mathbf{b})]$. Otherwise, the bias is larger than 0, indicating a non-zero penalty at pixel $\mathbf{m}$.

To understand how changing image scale affects the bias in smooth regions, we assume that the original continuous displacement field is *sufficiently sampled* by the image where the unknown displacements at sub-pixel level in a local region can be approximated by a monotone function between the neighboring pixels. This assumption is not restrictive and can be satisfied in many images.

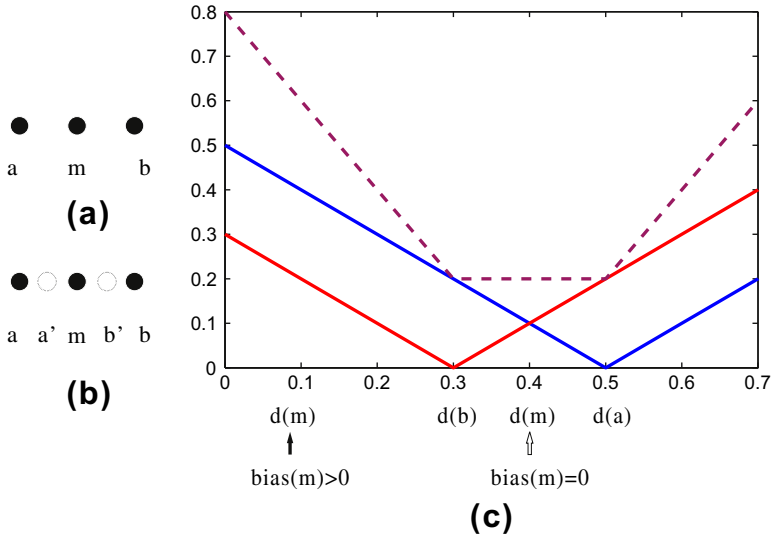

Fig. 2. Illustration of the regularization bias. (a) Three consecutive pixels in a row denoted by **a**, **m**, and **b** respectively. (b) Corresponding pixels in the high-resolution image. (c) Regularization energies for pixel **m**. The blue and red curves plot the smoothness costs for $|\mathbf{a}|$ and $|\mathbf{b}|$ respectively. The dashed curve represents the total regularization cost $\mathbf{d}_{reg}(\mathbf{m})$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Average errors (in terms of regularization bias) for different flow fields.

Reg. bias	$\overline{bias}(\mathbf{A})$	$\overline{bias_u}(\mathbf{A_u})$	$\overline{bias_u}(\mathbf{A})$
Avg.	0.0309	0.0089	0.0092

Claim 2. *If the displacement field in a region is sufficiently sampled (that is, the displacement change between neighboring samples are monotone) and the total variation regularizer is used, the regularization bias for a pixel in that region in the original image is larger than or equal to that in the ideal upsampled version.*

Proof. For simplicity's sake, we also prove it in 1D, while the same result can be derived in 2D. The regularization bias can be explicitly written as

$$bias(\mathbf{m}) = \begin{cases} 0 & \text{if}(\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a}))(\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b})) \leqslant 0 \\ \min(\|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a})\|, \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b})\|) & \text{otherwise} \end{cases}$$

where **a** and **b** are the neighbors of **m** as indicated in Fig. 2. In an ideal upsampled image, the neighbors of **m** are denoted as $\mathbf{a}'$ and $\mathbf{b}'$ (Fig. 2b). The bias $bias_u$ for the ideal upsampled image is given by

$$bias_u(\mathbf{m}) = \begin{cases} 0 & \text{if}(\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a}'))(\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b}')) \leqslant 0 \\ \min(\|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a}')\|, \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b}')\|) & \text{otherwise} \end{cases}$$

We show $bias_u(\mathbf{m}) \leqslant bias(\mathbf{m})$. First, if $bias(\mathbf{m}) = 0$, because

$$(\mathbf{d}(\mathbf{a}') - \mathbf{d}(\mathbf{a}))(\mathbf{d}(\mathbf{b}') - \mathbf{d}(\mathbf{b})) \leqslant (\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a}))(\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b})) \leqslant 0,$$

based on the monotone property, we get $bias_u(\mathbf{m}) = 0$. Second, if $bias(\mathbf{m}) \neq 0$, also according to $\|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a}')\| \leqslant \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{a})\|$ and $\|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b}')\| \leqslant \|\mathbf{d}(\mathbf{m}) - \mathbf{d}(\mathbf{b})\|$, we have $bias_u(\mathbf{m}) \leqslant bias(\mathbf{m})$. □

To validate the Claim 2 and test how the regularization bias changes in general images, which contain both high and low frequency structures, we conduct experiments on the Middlebury optical flow data set [2]. We downsample the given flow field with the factor 2 to simulate the low resolution field and evaluate the regularization bias on both levels of images. The bias for the 2D flow fields is computed using Eq. (19). Results are listed in Table 2. The average bias of all pixels in low resolution field is denoted by $\overline{bias}(\mathbf{A})$. For the high resolution field, we calculate the average bias (a) on all pixels of the high resolution grid (denoted by $\overline{bias_u}(\mathbf{A_u})$) and (b) on the same pixel set of the low resolution counterpart (denoted by $\overline{bias_u}(\mathbf{A})$) respectively. Both of them are much smaller than that on the low resolution grid. The bias decreases around 70% in average.

6. An upsampling framework

We have shown that by using an ideal upsampled image pair, the matching accuracy generally increases while the regularization bias works the other way around especially in smooth regions. When there is no high resolution input in stereo matching or optical flow estimation, we propose a general upsampling framework with only low-resolution images to improve sub-pixel matching accuracy. Our algorithm is sketched in Table 3. It has procedures to suppress possible errors introduced in upsampling.

Preprocess. We first upsample the input images L_0 and L_1. We tested representative upsampling methods, including the interpolation (bilinear and bicubic), reconstruction [22], and learning based methods [15].

Matching. Any matching algorithm can be used in our system, since our framework is a general one for improving sub-pixel correspondence. The matching algorithm is performed on the input and upsampled image pairs. It results in two displacement maps $\mathbf{d}_H$ and $\mathbf{d}_L$. We then downsample $\mathbf{d}_H$ to $\mathbf{d}_{L'}$ in order to construct a new map in its original resolution. The reason that we still keep $\mathbf{d}_L$ is that $\mathbf{d}_{L'}$ is not everywhere better than $\mathbf{d}_L$. Although the upsampled image matching scheme generally improves the result, it may also deteriorate a few areas (e.g., highly textured regions, or object boundaries with step edges). In the next step, we fuse these the displacement maps and select optimal values.

Fusing. Our fusing algorithm computes a binary map c that indicates which displacement value to use (from either $\mathbf{d}_L$ or $\mathbf{d}_{L'}$). We minimize energy E_c in order to infer c. E_c contains 3 terms and is written as

$$E_c = E_{c1} + E_{c3} + E_{c2}. \tag{22}$$

E_{c1} is the data term, defined as

Table 3
Overview of our method.

0. **Preprocess**:
   0.1 Upscale input images
1. **Matching**:
   1.1 Estimate the displacement field $\mathbf{d}_L$ with sub-pixel accuracy on the input image pair
   1.2 Estimate displacement map $\mathbf{d}_H$ on the upsampled images
   1.3 Downsample $\mathbf{d}_H$ to $\mathbf{d}_{L'}$
2. **Fusing**:
   2.1 Fuse $\mathbf{d}_{L'}$ and $\mathbf{d}_L$ based on multiple cues to reject outliers

$$E_{c1} = \sum_{\mathbf{m}} (1 - c_{\mathbf{m}}) \|\mathcal{M}_L(\mathbf{d}_L)\|^2 + c_{\mathbf{m}} \|\mathcal{M}_{L'}(\mathbf{d}_{L'})\|^2, \quad (23)$$

where $\mathcal{M}_L$ is the matching confidence given by

$$\mathcal{M}_L(\mathbf{d}) = L_1(\mathbf{m} + \mathbf{d}(\mathbf{m})) - L_0(\mathbf{m}), \quad (24)$$

with $\mathbf{m}$ being the pixel coordinate at which correspondence is computed. $\mathcal{M}_{L'}$ is defined in a similar way as Eq. (24). E_{c1} is a selection between two measures given the value of $c_{\mathbf{m}}$ being either 0 or 1. Minimizing E_{c1} favors small cost.

E_{c2} is the regularization term, defined as

$$E_{c2} = \sum_{\mathbf{m}} (\alpha c_{\mathbf{m}} \|\nabla L\|^2 + \beta \|c_{\mathbf{m}} - V_{\mathbf{m}}\|^2). \quad (25)$$

It consists of two energies that are balanced by weights α and β. Energy $c_{\mathbf{m}}\|\nabla L\|^2$ penalizes strong gradients where $c_{\mathbf{m}} = 1$ due to the consideration that the matching quality for regions with abrupt color transition is unknown. Energy $\|c_{\mathbf{m}} - V_{\mathbf{m}}\|^2$ is based on a heuristic that upscaled image matching only improves sub-pixel accuracy. $V_{\mathbf{m}}$ is expressed as

$$V(\mathbf{m}) = \begin{cases} 1 & \text{if} \|\mathbf{d}_L(\mathbf{m}) - \mathbf{d}_{L'}(\mathbf{m})\|_1 < 1 \\ 0 & \text{Otherwise} \end{cases}$$

It indicates if $\mathbf{d}_L$ and $\mathbf{d}_{L'}$ are quite different, the distance is very likely being caused by the error in the upscaling process. It is thus preferable to trust $\mathbf{d}_L$.

The final term is a spatial smoothness term E_{c3}:

$$E_{c3} = \sum_{\mathbf{m},\mathbf{n}\in\mathcal{N}(\mathbf{m})} \gamma \|c_{\mathbf{m}} - c_{\mathbf{n}}\|^2, \quad (26)$$

where γ is a weight and $\mathcal{N}(\mathbf{m})$ are the neighbors of $\mathbf{m}$.

Minimizing E_c is a binary optimization problem. We use graph cuts to solve it with $\alpha = 1$, $\beta = 10$, and $\gamma = 5$. The computed map c yields the final displacement map

$$\mathbf{d}_F(\mathbf{m}) = (1 - c(\mathbf{m}))\mathbf{d}_L(\mathbf{m}) + c(\mathbf{m})\mathbf{d}_{L'}(\mathbf{m}). \quad (27)$$

7. Implementation and experimental results

In this section, we first empirically evaluate a few state-of-the-art upsampling and super-resolution methods on the context of

Fig. 3. Image examples used to evaluate our approach.

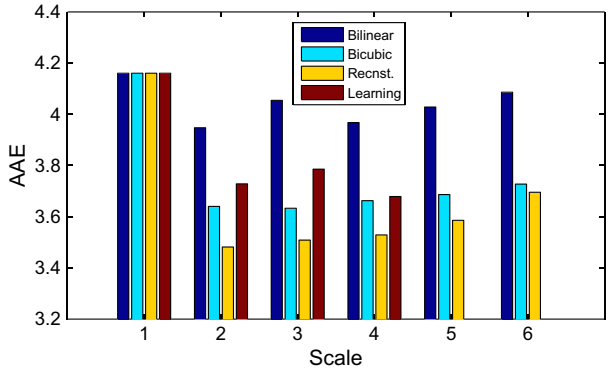

Fig. 4. AAEs of matching with different upscaling schemes.

Table 4
Average PSNRs on the test images for different methods.

Method	Factor 2	Factor 3	Factor 4
Nearest neighbor	26.48	24.56	22.13
Bilinear	26.51	24.68	22.87
Bicubic	26.96	24.83	22.93
Reconstruction-based	28.87	26.68	24.90
Example-based	27.97	24.83	23.94

Table 5
Average errors of the flow estimates on the original and upsampled images. Columns "*AAE*(3%)" and "*AAE*(5%)" list the errors where the input low-resolution images are with white Gaussian noise (standard deviation 3% and 5% respectively).

Method	AAE	EPE	AAE(3%)	AAE(5%)
L pair	4.11	0.16	6.13	7.27
HR pair	2.67	0.10	N/A	N/A
Nearest neighbor	4.67	0.18	6.17	7.17
Bilinear	3.32	0.13	5.60	6.73
Bicubic	3.19	0.12	5.58	7.03
Reconst.-based	3.09	0.11	5.53	7.16
Example-based	3.22	0.12	5.73	7.40

Table 6
AAE and EPE for the results computed on the 8 image pairs using different methods.

	Dime.		Grove2		Grove3		Hydrangea	
A. L	8.26	0.41	3.96	0.28	8.42	0.99	3.10	0.35
A. U	4.16	0.23	3.28	0.22	7.61	0.91	2.70	0.28
B. L	3.21	0.16	2.67	0.18	6.53	0.76	2.42	0.20
B. U	2.70	0.13	1.95	0.12	5.89	0.62	2.07	0.16
B.UD	3.07	0.15	2.27	0.17	6.19	0.66	2.24	0.20
B.LN	6.63	0.35	3.34	0.21	7.89	0.90	3.22	0.31
B.UN	5.47	0.28	2.79	0.18	7.18	0.82	2.74	0.24
	RWhale		Urban2		Urban3		Venus	
A. L	8.06	0.27	8.93	3.10	13.30	1.53	7.47	0.54
A. U	5.94	0.19	8.73	3.11	11.56	1.49	7.47	0.55
B. L	5.27	0.15	3.21	0.69	9.98	1.20	6.41	0.46
B. U	4.19	0.12	2.75	0.80	7.18	0.86	5.53	0.31
B.UD	4.83	0.15	3.19	0.79	8.79	1.19	5.79	0.43
B.LN	9.51	0.30	5.65	1.02	15.15	1.58	7.39	0.49
B.UN	8.27	0.25	5.87	1.05	14.88	1.48	6.91	0.47

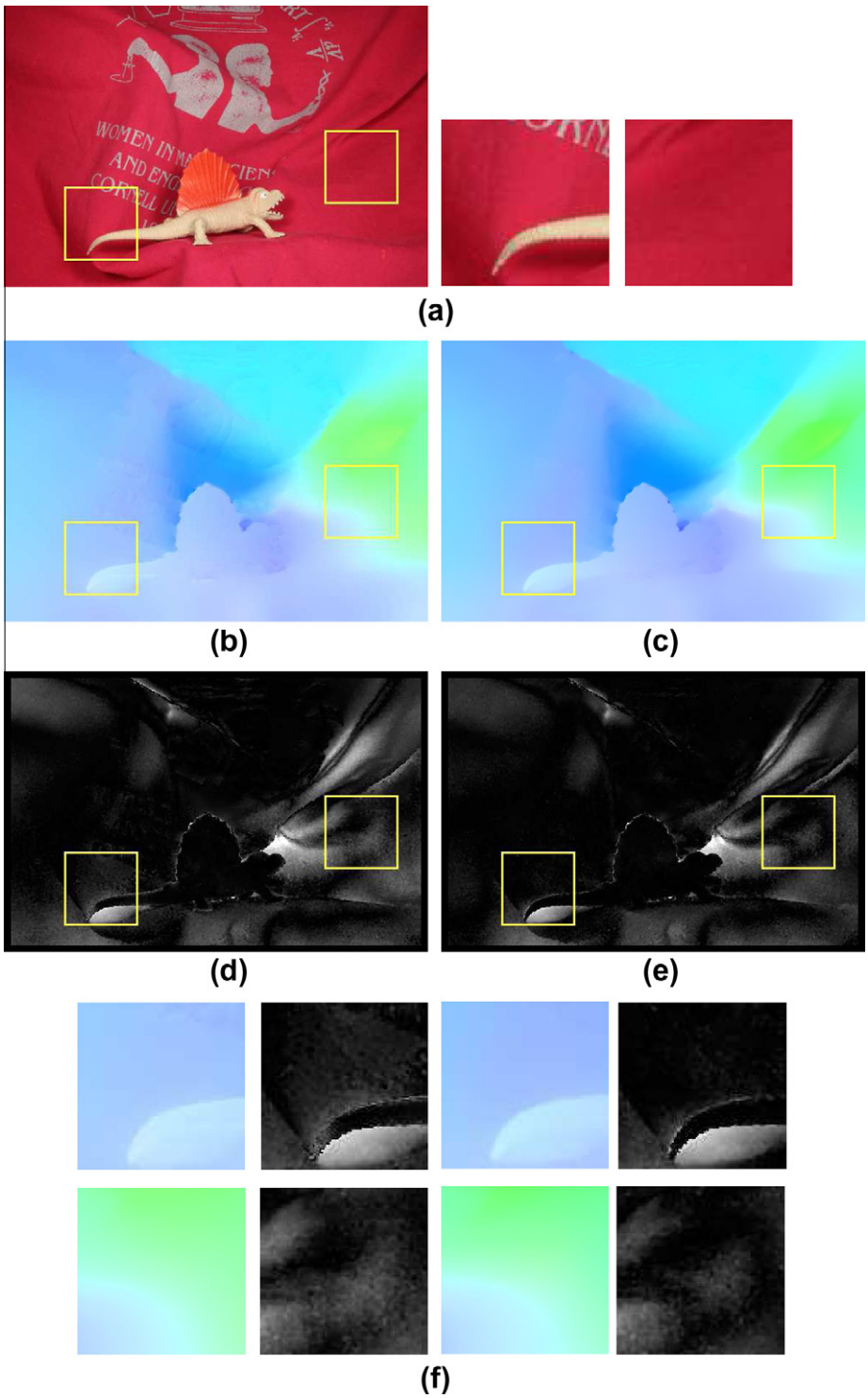

Fig. 5. Visual comparison of motion estimation results. (a) An input image with two highlighted patches. (b) Flow estimates in the original resolution. (c) Flow computed using our method. (d)–(e) Corresponding AAE maps. They are linearly scaled for visibility. (f) Close-ups.

general matching tasks and then apply our method to optical flow estimation and stereo matching.

7.1. Comparison of upsampling methods

To evaluate the upsampling methods, we ran two sets of experiments. In the first set, we evaluate how different methods, including interpolation (bilinear and bicubic), reconstruction [22], and learning based methods [15], affect the reconstruction error.

7.1.1. Evaluating upsampling quality

The image data are collected from the Kodak image dataset, and the Middlebury stereo and optical flow websites [21,2]. Sample images are shown in Fig. 3. We regard the collected images as the ground truth high-resolution images, and create low-resolu-

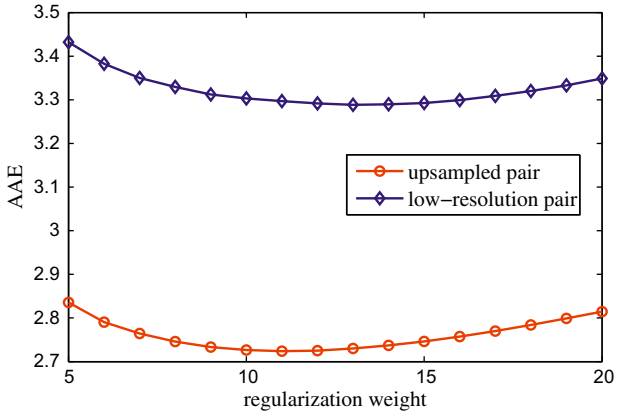

Fig. 6. AAEs for the "Dimetrodon" dataset with different regularization weights.

tion ones by downsampling them with scale factors 2, 3 and 4 respectively using bicubic interpolation. These coarse-scale images are upsampled with different methods listed in Table 4. The Peak Signal to Noise Ratios (PSNRs) between the upsampled and ground truth images are calculated to evaluate the quality.

Table 4 lists the PSNRs by different methods, all of which use default parameter settings. For the reconstruction- and learning-based methods, we use the implementation [22,15] of the authors. It is found that the reconstruction- and learning-based methods outperform interpolation-based resizing.

7.1.2. Matching quality evaluation

Our second set of experiments were conducted to measure how different upsampled images affect correspondence estimation. In this step, we implement the optical flow estimation method of Bruhn and Weickert [7] to compute motion using both the original and upsampled image pairs. Middlebury optical flow images (with the ground truth flow) are used to measure the flow accuracy with

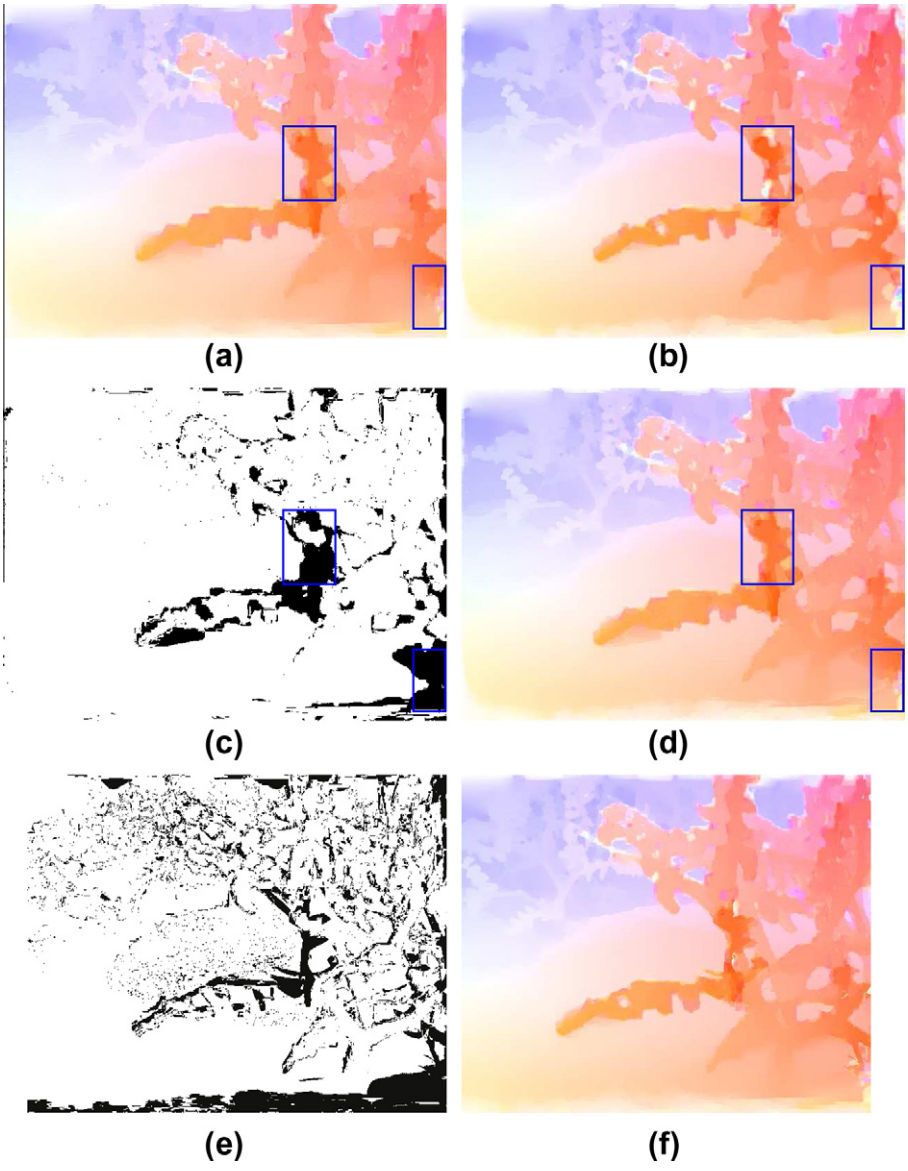

Fig. 7. Fusing. (a) Flow map computed on the low resolution image pair. (b) Flow estimated on the upsampled image pair. (c) The computed binary map *c*. (d) The final result after fusing. The errors near object boundaries are reduced. (e) Optimal binary map computed using AAEs as guidance. (f) Fusing result using (e). The AAEs for (a), (b), (d) and (f) are 8.42, 7.75, 7.61 and 7.05 respectively.

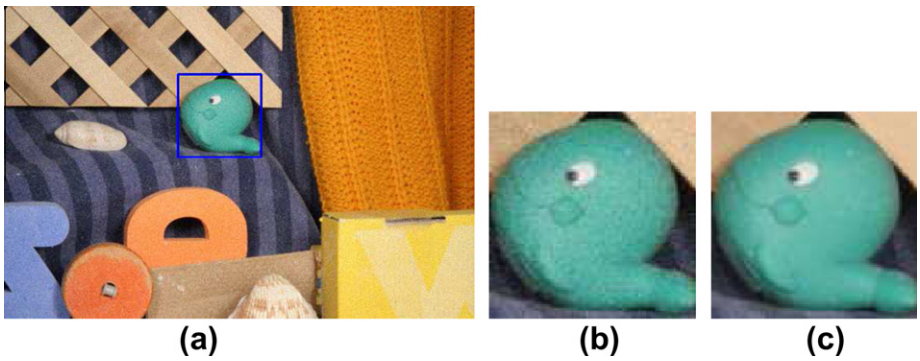

Fig. 8. A noisy image example. (a) One "RubberWhale" image with additive noise. (b) Close-up of (a). (c) The corresponding patch in the noise-free image.

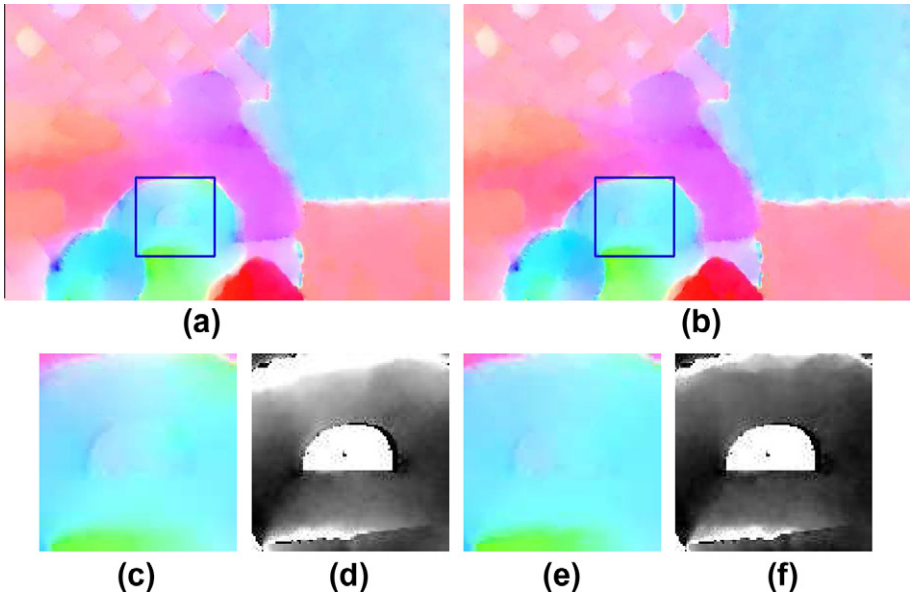

Fig. 9. Flow on a noisy example. (a) Flow estimates from the input image pair. (b) Flow map obtained by employing our upsampling framework. The AAE drops from 9.52 to 8.31.

regard to the Average Angular Error (AAE) and End Point Error (EPE) [2].

The detailed procedure is as follows. For the input image pairs with the ground truth optical flow, we first downsample them with scale factor 2. The coarse-level images are then upsampled using different methods. We estimate flow on these images. Table 5 (columns 2–3) lists the calculated errors. In the "L pair" and "HR pair" rows of Table 5, we show the average errors for the low and high-resolution images. "HR pair" has 35% less error than "L pair". Other rows contain errors calculated on images upsampled by different methods. They demonstrate that all tested methods can improve matching accuracy except for the nearest neighbor re-sampling because it cannot preserve image structure well.

We show in Fig. 4 the average matching errors (AAEs) with different scale factors using our upsampling-matching framework. It shows that better results are consistently yielded after upsampling the images.

7.2. Application to general optical flow estimation

Optical flow is usually computed either by directly minimizing the energy function using a continuous optimization method [4], or by solving the corresponding Euler–Lagrange equations of the following variational model [6,8]:

$$\int_{\Omega} \rho(\|L_1(\mathbf{x}+\mathbf{d}) - L_0(\mathbf{x})\|^2) + \lambda \cdot \rho(\|\nabla \mathbf{d}\|^2) d\mathbf{x}.$$

We experiment with two flow estimation algorithms using our upsampling framework. One is our implementation of Black and Anandan's method [4] where a Lorentzian ρ-function is adopted to reject outliers. The second estimator is based on our implementation of [8] using a non-linear multi-grid framework (Full Approximation Scheme). In this implementation, we use the Total Variation regularizer as the ρ-function. Reconstruction-based method [22] is adopted to upsample input images.

7.2.1. Error comparison

Table 6 shows the flow errors yielded from these two methods, denoted as 'A.' [4] and 'B.' [7] respectively. 'L' and 'U' denote flow estimates on the input images and using our approach (sketched in Table 3). 'LN' and 'UN' are their counterparts by adding 3% Gaussian noise. For flow evaluation, we remove 10-pixel-width image boundary because these pixels are not well handled [2]. We record an improvement up to 50% ('Dimetrodon' set) and 29% ('Grove' set) respectively for the methods of Black and Anandan (A.), and of Bruhn and Weickert (B.). The average errors are also reduced by 16% and 21% respectively. As our framework improves sub-pixel accuracy, robust approaches, such as 'B.', can be benefitted more. One

Fig. 10. Stereo matching examples. (a) Reference image in each image pair. (b) Occlusion map. (c) Disparity map obtained only using the low resolution images. (d) Disparity map estimated by our upsampling framework.

Table 7
Average disparity errors. The root mean square errors (RMSEs) from our fusing framework is consistently smaller than those obtained on low resolution images.

Method	Art	Dolls	Reindeer	Cloth3	Cones
BP. L	1.22	0.71	0.91	0.49	0.70
BP. U	1.18	0.51	0.83	0.31	0.63
BP. L ($\overline{O}$)	0.57	0.51	0.48	0.47	0.48
BP. U ($\overline{O}$)	0.47	0.39	0.39	0.26	0.33

visual comparison is given in Fig. 5 where the AAEs are 3.29 (in the original resolution) and 2.72 (using our method). Close-ups in (f) show better estimation can obtained especially for regions where the flow vectors change smoothly.

7.2.2. Comparison of costs

To further verify that both the data and smoothness terms estimation can be profitted in our framework, we experiment with only changing the data term in the optimization process and leaving the regularization term intact. Specifically, we upscale the input image pairs and compute the data term on them. Then we downsample the data costs with a box filter and combine them with the original regularization term to compute a flow field. This process only reduces the data interpolation errors. The final flow field is computed with our fusion step. Row "B.UD" in Table 6 contains the errors calculated with this configuration. Interestingly, the results have 11% improvement in average, which is about the half of the total improvement, which is 21%, achieved using our complete framework. It indicates that the upscaling strategy not only reduces data interpolation error, but also mitigates the side-effect of regularization, such as over-smoothing.

Note that our upscale-matching scheme is not equivalent to setting different regularization weights on the original images. We verify it by varying the weight λ in a range for both the input and upsampled image pairs. The graph in Fig. 6 shows that the variation of AAE by this means is not significant. For all weights we tested, the AAE obtained with our approach is always smaller.

7.2.3. Effect of fusion

We show in Fig. 7 an example to demonstrate the effectiveness of our fusion algorithm. (a) and (b) show flow maps $\mathbf{d}_L$ and $\mathbf{d}_{L'}$ respectively where the AAEs are 8.42 and 7.75. Although (b) is more accurate, there are problems near motion boundaries. Two such patches are highlighted. (c) is the extracted fusing map c by Eq. (22), where the pixels with incorrect flow are appropriately labeled. (d) shows the final result after fusing, with AAE reduced to 7.61. For comparison, we also compute the optimal fusing map (e) taking AAEs as a guidance – that is, only flow vectors with the smaller AAE are selected. The similarity of (c) and (e) verifies the effectiveness of our fusing step.

7.2.4. Effect of noise

We also evaluate the system stability against image noise. Fig. 8a shows an exemplar noisy input image. Rows tagged as 'LN' and 'UN' in Table 6 contain estimation errors on these images with 3% white noise. While error increases with noise, our upsampling framework still works. One visual comparison is shown in Fig. 9.

7.3. Application to stereo matching

We apply our method to two-view stereo matching [21] to show that this framework is also helpful when discrete optimization is employed. Experiments are conducted on the Middlebury stereo dataset [21,20]. To evaluate the performance in terms of sub-pixel correspondence accuracy, we downsample the obtained images by a factor of 5, which makes the scaled-down ground truth disparities have enough sub-pixel accuracy. Then our upsampling framework is applied to the low-resolution stereo images. Fig. 10a shows the input low resolution reference images.

The energy function (1) used here for stereo matching also combines the data and regularization terms, where the 1D disparities are discretized at 0.2-pixel and 0.4-pixel intervals for the low-resolution and upsampled images respectively. The numbers of disparity labels for the image pairs are the same.

Stereo matching using belief propagation (BP) [24] is adopted to compute the disparities. Since the disparity values in the occluded regions are ambiguous, we detected occluded pixels in the matching process. Fig. 10b shows the computed occlusion map where disparities are only evaluated in gray regions.

Fig. 10c and d respectively show the computed disparity maps from the low resolution images and from the upsampled version. The disparity quality in (d) is higher than that in (c), especially in smooth regions. We show in Table 7 the average disparity errors (RMSEs), with and without including occlusion pixels. It shows that the estimates in non-occluded regions get improved a lot. Disparity errors in occluded regions are simply unpredictable and thus cannot be handled well.

8. Discussion

The proposed upsampling-correspondence scheme is a simple and general framework suitable for matching-based computer vision methods to improve computation accuracy, especially in the subpixel level. Although it requires extra computation, compared to the conventional difficulty in subpixel correspondence establishment, this scheme finds many applications given the development of both the matching algorithms and computing hardware. Presently, the fast optical flow algorithms only take seconds to compute a flow field with image size 640×480. Comparably, increasing the sub-pixel matching accuracy by 10% is still considered challenging and significant. This paper shows that our approach can yield more than 20% improvement.

The effect of image noise is considered and evaluated in our experiments for different upsampling methods. The last two columns of Table 5 (AAE(3%) and AAE(5%)) show that the results using the upsampled images are still better than those obtained on the input images (up to an improvement of 10%).

9. Concluding remarks

In this paper, we studied the sub-pixel correspondence problem by analyzing both the data term interpolation errors and the regularization bias. We show that for natural images with locally smooth-varying regions, the interpolation error from a high-resolution image is smaller than that produced on its low-resolution version. In addition, we analyzed how change in image scales affects the regularization bias and how this analysis justifies the use of an upsampling framework. A general framework for improving sub-pixel matching accuracy was proposed and was tested on optical flow estimation and stereo matching. Although upsampling requires extra computation, we observe consistent and significant quality improvement, which would be difficult to obtain otherwise.

Acknowledgments

The authors would like to thank the associate editor and all the anonymous reviewers for their time and effort. This work is supported by a grant from the Research Grants Council of the Hong Kong SAR (Project No. 413110) and by NSF of China (key project No. 61133009)

References

[1] Simon Baker, Takeo Kanade, Super-resolution optical flow, Technical Report CMU-RI-TR-99-36, The Robotics Institute, Carnegie Mellon University, 1999.
[2] Simon Baker, Daniel Scharstein, J.P. Lewis, Stefan Roth, Michael J. Black, Richard Szeliski, A database and evaluation methodology for optical flow, in: ICCV, 2007, <http://www.vision.middlebury.edu/flow/eval/results/results-a1.html>.
[3] Stan Birchfield, Carlo Tomasi, A pixel dissimilarity measure that is insensitive to image sampling, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (4) (1998) 401–406.
[4] Michael J. Black, P. Anandan, The robust estimation of multiple motions: parametric and piecewise-smooth flow fields, Computer Vision and Image Understanding 63 (1) (1996) 75–104.
[5] Jean-Yves Bouguet, Pyramidal implementation of the lucas kanade feature tracker: description of the algorithm, Opencv document, Intel Corporation, Microprocessor Research Labs, 2000.
[6] Thomas Brox, Andrés Bruhn, Nils Papenberg, Joachim Weickert, High accuracy optical flow estimation based on a theory for warping, in: ECCV (4), 2004, pp. 25–36.
[7] Andrés Bruhn, Joachim Weickert, Towards ultimate motion estimation: Combining highest accuracy with real-time performance, in: ICCV, 2005, pp. 749–755.
[8] Andrés Bruhn, Joachim Weickert, Christoph Schnörr, Lucas/kanade meets horn/schunck: combining local and global optic flow methods, International Journal of Computer Vision 61 (3) (2005) 211–231.
[9] R.B. Fisher, D.K. Naidu, A comparison of algorithms for subpixel peak detection, Image Technology, Advances in Image Processing, Multimedia and Machine Vision (1996) 385–404.
[10] Rik Fransens, Christoph Strecha, Luc J. Van Gool, Optical flow based super-resolution: a probabilistic approach, Computer Vision and Image Understanding 106 (1) (2007) 106–115.
[11] Stefan K, Gehrig and Uwe Franke, improving stereo sub-pixel accuracy for long range stereo, in: ICCV VRML Workshop, 2007, pp. 1–7.
[12] Russell C. Hardie, Kenneth J. Barnard, Ernest E. Armstrong, Joint map registration and high-resolution image estimation using a sequence of undersampled images, IEEE Transactions on Image Processing 6 (12) (1997) 1621–1633.
[13] Li Hong, George Chen, Segment-based stereo matching using graph cuts, in: CVPR (1), 2004, pp. 74–81.
[14] Berthold K.P. Horn, Brian G. Schunck, Determining optical flow, Artificial Intelligence 17 (1–3) (1981) 185–203.
[15] Kwang In Kim, Younghee Kwon, Example-based learning for single-image super-resolution, in: DAGM-Symposium, 2008, pp. 456–465.
[16] Vladimir Kolmogorov, Ramin Zabih, Computing visual correspondence with occlusions via graph cuts, in: ICCV, 2001, pp. 508–515.

[17] Victor S. Lempitsky, Stefan Roth, Carsten Rother, Fusionflow: discrete-continuous optimization for optical flow estimation, in: CVPR, 2008.
[18] Diego Nehab, Szymon Rusinkiewicz, James Davis, Improved sub-pixel stereo correspondences through symmetric refinement, in: ICCV, 2005, pp. 557–563.
[19] Emmanouil Z. Psarakis, Georgios D. Evangelidis, An enhanced correlation-based method for stereo correspondence with sub-pixel accuracy, in: ICCV, 2005, pp. 907–912.
[20] Daniel Scharstein, Chris Pal, Learning conditional random fields for stereo, in: CVPR, 2007.
[21] Daniel Scharstein, Richard Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision 47 (1–3) (2002) 7–42.
[22] Qi Shan, Zhaorong Li, Jiaya Jia, Chi-Keung Tang, Fast image/video upsampling, ACM Transactions on Graphics 27 (5) (2008) 153:1–153:7.
[23] Masao Shimizu, Masatoshi Okutomi, Precise sub-pixel estimation on area-based matching, in: ICCV, 2001, pp. 90–97.
[24] Jian Sun, Nan-Ning Zheng, Heung-Yeung Shum, Stereo matching using belief propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (7) (2003) 787–800.
[25] Richard Szeliski, Daniel Scharstein, Sampling the disparity space image, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (3) (2004) 419–425.
[26] Q. Tian, M.N. Huhns, Algorithms for subpixel registration, Computer Vision, Graphics, and Image Processing 35 (2) (1986) 220–233.
[27] Werner Trobin, Thomas Pock, Daniel Cremers, Horst Bischof, An unbiased second-order prior for high-accuracy motion estimation, in: DAGM-Symposium, 2008, pp. 396–405.
[28] Oliver J. Woodford, Philip H.S. Torr, Ian D. Reid, Andrew W. Fitzgibbon, Global stereo reconstruction under second order smoothness priors, in: CVPR, 2008.
[29] Jiangjian Xiao, Hui Cheng, Harpreet S. Sawhney, Cen Rao, Michael A. Isnardi, Bilateral filtering-based optical flow estimation with occlusion detection, in: ECCV (1), 2006, pp. 211–224.
[30] Li Xu, Jianing Chen, Jiaya Jia, A segmentation based variational model for accurate optical flow estimation, in: ECCV (1), 2008, pp. 671–684.
[31] Qingxiong Yang, Liang Wang, Ruigang Yang, Henrik Stewénius, David Nistér, Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling, in: CVPR (2), 2006, pp. 2347–2354.
[32] Wen-Yi Zhao, Harpreet S. Sawhney, Is super-resolution with optical flow feasible? in: ECCV (1), 2002, pp. 599–613.
[33] C. Lawrence Zitnick, Nebojsa Jojic, Sing Bing Kang, Consistent segmentation for optical flow estimation, in: ICCV, 2005, pp. 1308–1315.