Joint Estimation in Stereo Super-resolution

Arnav V. Bhavsar
Image Processing and Computer Vision Lab
IIT Madras
Email: arnav_bhavsar@yahoo.com

A.N. Rajagopalan
Image Processing and Computer Vision Lab
IIT Madras
Email: raju@ee.iitm.ac.in

***Abstract*—This paper considers the motion based super-resolution problem in the stereo vision domain. For stereo images the registration information at the high resolution (HR) is same as the disparity between the images at HR. We point out that registration information cannot be mapped from low resolution (LR) coordinate system to HR coordinate system in cases where shift of the pixels is not constant (e.g. for images from a stereo setup). Because of such a 'pixel migration' phenomenon, an *a priori* process of computing registration information at HR, as is popular in super-resolution techniques, cannot be followed. We hence use a MAP-MRF framework to simultaneously estimate the HR disparity map and the super-resolved (SR) image given low resolution images from a stereo setup. In this framework, both the SR image and HR disparity are modeled as MRF.**

***Index Terms*—Stereo, super-resolution, MAP-MRF, disparity estimation.**

I. Introduction

Super-resolution techniques exploit either of the various visual cues such as motion [1], [2], [3], [4], [5], [6], blur [7], [8], [9], photometry [10] etc or follow a learning based approach [11], [12]. All of these involve estimating a HR image from multiple LR observations each of which ideally contains some unique information about the HR image. This work falls into the category of motion based super-resolution in which the LR observations are relatively shifted and down-sampled versions of the original HR image.

Motion based super-resolution has progressed from simple cases of translational motion in planar scenes [1] to multi parametric motion [3], [4] and optical flow [13], [14]. An important step in motion based super-resolution is the computation of registration information in the HR coordinate system. The registration estimates for planar scenes involving two parameter translational motion can be computed as a pre-processing step and then fed into the main super-resolution algorithm. More sophisticated Bayesian techniques simultaneously compute the registration parameters along with the super-resolved image [3], [4]. However, when the motion is not parametric (e.g. scenes with moving objects), the problem of estimating the registration information essentially becomes a combinatorial optimization with large number of unknowns. Recently, optical flow super-resolution methods have been proposed that iterate over estimating the high-resolution dense optical flow between and the super-resolved image [13], [14].

In this paper, we concern ourselves with a stereo setup to capture the LR images in a static 3d scene. The relative shift of the pixels, between images, is not global due to the 3D nature of the scene. This non-global shift of pixels, what we term as 'pixel migration' phenomenon, does not allow *a priori* mapping of the shifts from the LR to HR coordinates, as is the traditional practice in super-resolution approaches for planar scenes. The shifts or the registration information required for SR is nothing but the HR disparity in case of a stereo setup. Hence, in this work, we solve the problem of simultaneously estimating the HR disparity and the SR image from LR stereo observations. This problem can also be looked at from a stereo vision point of view of estimating the high-resolution disparity from LR stereo observations. The novelty of the work lies in the integration of the super-resolution and the disparity estimation problems in a single framework. We pose the problem as a MAP-MRF estimation which essentially leads to a energy minimization problem with prior knowledge. We model both the disparity and the image priors as MRF. Apart from allowing a large class of convex and non-convex priors, we show that the MRF modeling of priors also makes the posterior have the 'locality' property. Since the energy is minimized for disparity and image in each iteration, a practical advantage of the theorem is that energies can be computed locally rather than on the whole HR grid in each iteration. This drastically reduces the computation complexity of techniques like simulated annealing or ICM which are known to perform well in super-resolution problems [15].

The paper is organized as follows. The next section discusses the image formation and the necessity of simultaneous estimation of HR disparity and SR image in a stereo super-resolution problem. Section 3 covers the joint MAP-MRF estimation of HR disparity and SR image. We provide the experimental results in section 4 and conclude in section 5. In this paper, we consider a simple case of translational camera motion that is used commonly in disparity estimation problems.

II. Image formation: Pixel migration

Typically, in motion based super-resolution we are given low resolution observations $[y_1, y_2, ..., y_N]$ of size $N_1 \times N_2$. These low resolution observations are modeled to be generated when a high resolution image x of size $L_1 \times L_2$ is warped, blurred and down-sampled; the down-sampling factor being $\frac{L_1}{N_1} \times \frac{L_2}{N_2}$. The warping is due to camera motion, the blurring is due to the camera point-spread function (PSF). This can be expressed in vector form as,

$$\mathbf{y}_i = DH_iW_i\mathbf{x} + \eta_i \quad (1)$$

Here, $\mathbf{y_i}$ is the lexicographically arranged i^{th} LR observation and D, H_i and W_i are the respectively the down-sampling, blur and warp matrices that produce $\mathbf{y}_i$ from the HR image $\mathbf{X}$. Without loss of generality, for this particular work, we assume the camera to be a pinhole and hence the H_i matrices are essentially I. Hence the above equation can be expressed in scalar form as,

$$y_i(n_1,n_2) = \sum_{l_1,l_2=1}^{L_1,L_2} d(n_1,n_2,l_1,l_2)\cdot x(\theta_{1i}(l_1),\theta_{2i}(l_2)) + \eta_i(n_1,n_2) \quad (2)$$

where, $d(n_1,n_2,l_1,l_2)$ is the element of the D matrix that maps the $(l_1,l_2)^{th}$ pixel in HR image $x(\theta_{1i}(l_1),\theta_{2i}(l_2))$ to the $(n_1,n_2)^{th}$ pixel in the i^{th} LR image. The transformations θ_{1i} and θ_{2i} are the warping transformations that are encoded in the matrix W_i. Clearly for the reference image $(i = 1)$, $\theta_{11}(l_1) = l_1$ and $\theta_{21}(l_2) = l_2$. For the stereo setup, the above equation can be rewritten as,

$$y_i(n_1,n_2) = \sum_{l_1,l_2=1}^{L_1,L_2} d(n_1,n_2,l_1,l_2)\cdot x(l_1-\delta_i(l_1),l_2-\delta_i(l_2)) + \eta_i(n_1,n_2) \quad (3)$$

$\delta_i(l_1)$ and $\delta_i(l_2)$ being the disparities in the x and y directions respectively. We now justify the necessity for simultaneous estimation of HR disparity and SR image. The D in equation (1) is considered to transform the HR image to LR observation by an averaging process of pixels. e.g. for a down-sampling factor of 2, 4 pixels in the HR image will be averaged to produce 1 pixel in the LR image. In a simple case global translation (for planar scenes), all the pixels shift by the same amount. A two step approach for motion estimation can be used in such situations where the translation is estimated at LR and then multiplied by the resolution factor K. This approach is valid in a global translation scenario because when the LR pixels are mapped to their corresponding HR 'averaging groups', each pixel in one such group is separated from a pixel in another such group by a constant amount which is equal to K times the distance between the corresponding LR pixels. When we have a situation where different pixels at HR move differently, there is no global translation. We term this non-constant pixel motion as 'pixel migration'. The motion (disparity) found at LR will be on a N_1N_2 sized grid. First of all finding accurate disparities or flow is in itself an ill-posed problem. Assuming that we are given the exact motion at LR, to map this motion at HR on a L_1L_2 sized grid, one has to interpolate and multiply the flow or disparity by K. Due to 'pixel migration', the distance between the pixels at HR can change in an arbitrary fashion unlike in the case of translation in a planar scene. This can result in an arbitrary formation of the HR 'averaging groups' in the shifted versions of the reference HR image. The registration estimate generated from a two step process of mapping the disparity estimates from a N_1N_2 grid to the L_1L_2 grid by interpolation and multiplication by the K is a very crude approximation of the true HR motion since, the 'averaging group' that we get at HR by doing this can be grossly different from the actual 'averaging group' which is getting averaged from the HR image to form the LR observation. Hence due to the 'pixel migration' phenomenon, a simultaneous estimation of HR disparity and SR image is inevitable.

In this work we restrict ourselves to binocular stereo images. Ideally, for our experiments, this will super-resolve the image in one direction and in the other direction we shall have Bayesian interpolation. However, in general, given the calibrated stereo images both x and y directions, we can achieve super-resolution in all the directions without increasing the complexity. This is because the number of disparity and image unknowns will still remain the same L_1L_2 irrespective of the number of images due to the calibrated nature of the stereo setup.

III. MAP-MRF ESTIMATION

We now express the problem of estimating the HR disparity map and the SR image using a MAP framework. More specifically, given the observations $\mathbf{y_1}$, $\mathbf{y_2}$, ... $\mathbf{y_n}$ from a stereo setup, we wish to solve for the HR disparities δ and the SR image $\mathbf{x}$. For simplicity of notation, we do away with the subscripts for δ that are used in the last section. This also supports the facts that the choice of reference disparities is essentially arbitrary. Let $\mathbf{Y_1}$, $\mathbf{Y_2}$, ... $\mathbf{Y_n}$ be the random fields associated with the observations $\mathbf{y_1}$, $\mathbf{y_2}$, ... $\mathbf{y_n}$ and let $\mathbf{\Delta}$, $\mathbf{X}$ be the random fields associated with the HR disparity δ and SR image $\mathbf{x}$. We wish to compute estimates $\widehat{\mathbf{\Delta}}$ and $\widehat{\mathbf{x}}$ such that,

$$\widehat{\delta},\widehat{\mathbf{x}} = \max_{\delta,\mathbf{x}} P(\mathbf{\Delta}=\delta,\mathbf{X}=\mathbf{x}|\mathbf{Y_1}=\mathbf{y_1}...\mathbf{Y_n}=\mathbf{y_n}) \quad (4)$$

Now since the SR image is consists of intensity values and the HR disparity is essentially a function of depth, these two are statistically independent. Hence using Bayes rule, and imposing statistically independence between SR image $\mathbf{X}$ and HR disparity $\mathbf{\Delta}$, the above equation can be written as,

$$\widehat{\delta},\widehat{\mathbf{x}} = \max_{\delta,\mathbf{x}} P(\mathbf{Y_1}=\mathbf{y_1}...\mathbf{Y_n}=\mathbf{y_n}|\mathbf{\Delta}=\delta,\mathbf{X}=\mathbf{x}) P(\mathbf{\Delta}=\delta)P(\mathbf{X}=\mathbf{x}) \quad (5)$$

As mentioned above, we consider a special case of a stereo pair as observation. Hence in our case, this can be re-written as

$$\widehat{\delta},\widehat{\mathbf{x}} = \max_{\delta,\mathbf{x}} P(\mathbf{Y_1}=\mathbf{y_1},\mathbf{Y_2}=\mathbf{y_2}|\mathbf{\Delta}=\delta,\mathbf{X}=\mathbf{x}) P(\mathbf{\Delta}=\delta)P(\mathbf{X}=\mathbf{x}) \quad (6)$$

The first term in the product on the right hand side of the equation is the likelihood term, that arises from the image formation model. From equation (1), assuming a pin hole camera and considering η be AWGN with variance σ^2

$$P(\mathbf{Y_1}=\mathbf{y_1},\mathbf{Y_2}=\mathbf{y_2}|\mathbf{\Delta}=\delta,\mathbf{X}=\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{N_1N_2}} \exp\left(-\sum_{i=1}^{2}\frac{\|\mathbf{y}_i - DW_i\mathbf{x}\|^2}{2\sigma^2}\right) \quad (7)$$

Here W contains the information about the disparities δ.

The Prior probabilities for disparity $P(\mathbf{\Delta} = \delta)$ and image $P(\mathbf{X} = \mathbf{x})$ are both modeled as MRFs. Such a 'double MRF' modeling in justified because in stereo vision works, it is well known that the disparity varies smoothly [18] and similarly in super-resolution applications, MRF modeling for images have shown good results [1], [15], [8]. Thus, we have,

$$P(\mathbf{\Delta} = \delta) = K \exp\left(-\sum_{c \in C_\delta} V_c^\delta(\delta)\right)$$
$$P(\mathbf{X} = \mathbf{x}) = K \exp\left(-\sum_{c \in C_\mathbf{x}} V_c^\mathbf{x}(\mathbf{x})\right) \quad (8)$$

where $V_c^x(x)$ and $V_c^\delta(\delta)$ are the clique potential functions for the image and depth respectively. Thus from equations (7) and (8), we can rewrite equation (6) as,

$$\widehat{\delta}, \widehat{\mathbf{x}} = \min_{\delta, \mathbf{x}} \left(\sum_{i=1}^{2} \frac{\|\mathbf{y}_i - DW_i\mathbf{x}\|^2}{2\sigma^2} + \sum_{c \in C_\delta} V_c^\delta(\delta) + \sum_{c \in C_\mathbf{x}} V_c^\mathbf{x}(\mathbf{x}) \right) \quad (9)$$

Thus, MAP estimation leads to an energy minimization problem, where the cost within the brackets in the above equation is to be minimized. The minimization can be performed in an iterative fashion where in one iteration the current estimate of image is used for estimating the disparity in the next iteration, the current estimate of disparity is fixed to estimate the image. The type of smoothness functions depend on the exact form of $V_c^x(x)$ and $V_c^\delta(\delta)$ which include various convex, piecewise constant, piecewise smooth, discontinuity adaptive functions.

The minimization of MAP-MRF energy functions as in (9) has gained a lot of popularity in recent years. Lately efficient graph based techniques, that update pixels in a parallel fashion, have been shown to perform very well in minimizing such energy functions [19], [20]. Although such techniques have been proved to be very successful for regular energy functions [20], their potential for non-regular functions, for cases where pixel averaging takes place such as super-resolution, deblurring etc, is not yet fully explored [21], [22]. On the other hand sequential updating methods such as simulated annealing (SA) or iterated conditional modes (ICM) have been shown to perform very well for super-resolution. Since our main stress is not on the solution approach but on the novelty of joint computation philosophy of SR image and HR disparity, we resort to SA for the energy minimization as given in (9).

The drawback of SA is its computationally complexity. However, MRF modeling of priors allows the posterior distribution to have the locality property. i.e. the posterior distribution is also an MRF. This property has a practical significance that it allows one to immensely reduce the complexity of sequential methods such as SA. In [15], the authors proved the locality property of posterior distribution for motion-based image super resolution in presence of space-variant blur. We now prove the locality property for our case of posterior distribution in simultaneous estimation of HR disparity and SR image.

Theorem 1: (i) For observations $\mathbf{y_1}, \mathbf{y_2}$, the posterior probability $P(\mathbf{\Delta} = \delta, \mathbf{X} = \mathbf{x} | \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2})$ has a Gibbs distribution with energy function,

$$U^{pos}(\mathbf{x}, \delta) = \left(\sum_{i=1}^{2} \frac{\|\mathbf{y}_i - DW_i\mathbf{x}\|^2}{2\sigma^2} + \sum_{c \in C_\delta} V_c^\delta(\delta) + \sum_{c \in C_\mathbf{x}} V_c^\mathbf{x}(\mathbf{x}) \right) \quad (10)$$

(ii) The conditional posterior probability is Markov and can be expressed as,

$$P[X_{p,q} = x_{p,q}, \Delta_{p,q} = \delta_{p,q} | X_{k,l} = x_{k,l}, \Delta_{k,l} = \delta_{k,l}; 1 \le (k,l) \le L_1 L_2, (k,l) \ne (p,q); \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}] = \frac{\exp(-U^{cpos}(\mathbf{x}, \delta))}{\sum_{\mathbf{x}_{p,q}, \delta_{p,q}} \exp(-U^{cpos}(\mathbf{x}, \delta))} \quad (11)$$

where the conditional posterior energy function is,

$$U^{cpos}(\mathbf{x}, \delta) = \sum_{i=1}^{2} \sum_{S - S_i} \frac{\|\mathbf{y}_i - DW_i\mathbf{x}\|^2}{2\sigma^2} + \sum_{c \in C, (p,q) \in c} V_c^\mathbf{x}(\mathbf{x}) + \sum_{c \in C, (p,q) \in c} V_c^\delta(\delta) \quad (12)$$

(iii) The posterior neighbourhood corresponding to site (p, q) is given by,

$$\psi_{p,q}^{pos} = \left(\psi_{p,q}^x \bigcup \psi_{p,q}^\delta\right) \bigcup_{i=1}^{2} \left(\bigcup_{(k,l) \in S - S_i} \zeta_{k,l}^{y_i} \right) \quad (13)$$

where $\psi_{p,q}^x$ and $\psi_{p,q}^\delta$ are the original neighbourhood corresponding to the basic MRF in the image and the disparity, $\zeta_{k,l}^{y_i}$ is the neighbourhood in the HR grid corresponding to the site (k, l) in the observation y_i, $S = \{(k,l) : 1 \le (k,l) \le N_1 N_2\}$ and $S_i = \left\{(k,l) : (i,j) \notin \zeta_{k,l}^{y_i}\right\}$.

Proof: From equations (4) to (8), the posterior probability can be written as,

$$P(\mathbf{\Delta} = \delta, \mathbf{X} = \mathbf{x} | \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}) = K \exp(-U^{pos}(\mathbf{x}, \delta)) \quad (14)$$

This is nothing but a Gibbs distribution where K is a normalizing constant and posterior energy function $U^{pos}(\mathbf{x}, \delta)$ is same as the bracketed term in equation (9) and (10). At this stage, we note that warping and down-sampling operations depend on a specific neighbourhood in the HR grid. That is, they have a finite local support of the order of a few pixels.

The conditional probability of $X_{p,q}$ and $\Delta_{p,q}$, given the observations is

$$P[X_{p,q} = x_{p,q}, \Delta_{p,q} = \delta_{p,q}; 1 \le (p,q) \le L_1 L_2 | \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}]$$
$$= P[X_{p,q} = x_{p,q}, \Delta_{p,q} = \delta_{p,q} | X_{k,l} = x_{k,l}, \Delta_{k,l} = \delta_{k,l}; 1 \le (k,l) \le L_1 L_2, (k,l) \ne (p,q); \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}] \cdot$$
$$P[X_{k,l} = x_{k,l}, \Delta_{k,l} = \delta_{k,l}; 1 \le (k,l) \le L_1 L_2, (k,l) \ne (p,q) | \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}] \quad (15)$$

Using equation (14) we have,

$$P[X_{p,q} = x_{p,q}, \Delta_{p,q} = \delta_{p,q} | X_{k,l} = x_{k,l}, \Delta_{k,l} = \delta_{k,l}; \\ 1 \leq (k,l) \leq L_1 L_2, (k,l) \neq (p,q); \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}] \\ = \frac{\exp(-U^{pos}(\mathbf{x}, \delta))}{\sum_{x_{p,q}, \delta_{p,q}} \exp(-U^{pos}(\mathbf{x}, \delta))} \quad (16)$$

where the summation in the denominator is over all possible levels of $x_{p,q}$ and $\delta_{p,q}$. We now define vectors,

$$\nu^i = \frac{1}{\sqrt{2}\sigma}(\mathbf{y}_i - DW_i\mathbf{x}) \quad (17)$$

Hence, the posterior energy function can be written as,

$$U^{pos}(\mathbf{x}, \delta) = \sum_{i=1}^{2} \sum_{1 \leq (k,l) \leq N_1 N_2} (\nu^i_{k,l})^2 + \sum_{c \in C_\delta} V_c^\delta(\delta) + \sum_{c \in C_\mathbf{x}} V_c^\mathbf{x}(\mathbf{x}) \quad (18)$$

The above equation can be decomposed as,

$$U^{pos}(\mathbf{x}, \delta) = \sum_{i=1}^{2} \sum_{S - S_i} (\nu^i_{k,l})^2 + \sum_{i=1}^{2} \sum_{S_i} (\nu^i_{k,l})^2 \\ + \sum_{c \in C_\delta, (p,q) \in c} V_c^\delta(\delta) + \sum_{c \in C_\delta, (p,q) \notin c} V_c^\delta(\delta) \\ + \sum_{c \in C_\mathbf{x}, (p,q) \in c} V_c^\mathbf{x}(\mathbf{x}) + \sum_{c \in C_\mathbf{x}, (p,q) \notin c} V_c^\mathbf{x}(\mathbf{x}) \quad (19)$$

Substituting equation (19) in equation (16) and canceling common terms in numerator and denominator, we obtain the conditional posterior as,

$$P[X_{p,q} = x_{p,q}, \delta_{p,q} = \delta_{p,q} | X_{k,l} = x_{k,l}, \delta_{k,l} = \delta_{k,l}; \\ 1 \leq (k,l) \leq L_1 L_2, (k,l) \neq (p,q); \mathbf{Y_1} = \mathbf{y_1}, \mathbf{Y_2} = \mathbf{y_2}] \quad (20) \\ = \frac{\exp(-U^{cpos}(\mathbf{x}, \delta))}{\sum_{x_{p,q}, \delta_{p,q}} \exp(-U^{cpos}(\mathbf{x}, \delta))}$$

where

$$U^{cpos}(\mathbf{x}, \delta) = \sum_{i=1}^{2} \sum_{S - S_i} (\nu^i_{k,l})^2 + \sum_{c \in C_\delta, (p,q) \in c} V_c^\delta(\delta) + \sum_{c \in C_\mathbf{x} (p,q) \in c} V_c^\mathbf{x}(\mathbf{x}) \quad (21)$$

It is important to note that in the first term, the inner sum is over a small area $S - S_i$ as opposed to the complete image in equation (18). The set $S - S_i$ consists of all pixels (k,l) in y_i such that $S - S_i = (k,l) : (p,q) \in \zeta^{y_i}_{k,l}$. This means that each (k,l) in the LR image y_i has a small neighbourhood $\zeta^{y_i}_{k,l}$ in the HR image. Hence, the posterior neighbourhood corresponding to site (p,q) is given by,

$$\psi^{pos}_{p,q} = \left(\psi^x_{p,q} \bigcup \psi^\delta_{p,q}\right) \bigcup_{i=1}^{2} \left(\bigcup_{(k,l) \in S - S_i} \zeta^{y_i}_{k,l} \right) \quad (22)$$

as required in equation (13). ■

The theorem implies that, in sequential methods, as one updates one pixel at a time in one iteration, the update in energy due to that pixel is computed only by using a small neighbourhood of that pixel rather than the whole image. This reduces the computation drastically, since this neighbourhood is much smaller than the image. Applying the above theorem, we perform the minimization in (9) by SA, iteratively for the SR image and HR disparity. More specifically, in the same metro-polis iteration, for a particular site in the HR grid, given the current estimate of the HR image we compute the current estimate of HR disparity and vice-versa. The parameters for the SA algorithm are empirically selected.

IV. Experiments and Results

In our experiments we used the stereo datasets from Middlebury [16], [18] and CMU [17]. We formed the LR stereo observations, down-sampled by 2 by averaging groups of 4 pixels in each image. The averaging had equal weight of 0.25 for each of the four pixels. We note here that the first term in the bracketed expression in equation (9) is what is commonly known as the data term in MAP-MRF minimization framework [19], [20]. In equation (9) this data term computes the squared distance or the L_2-norm since we assume the noise is AWGN. However, more generally we can replace it with L_1-norm or any other sensible distance metric [23]. We experimented with both L_1- and L_2-norm (i.e. squared and absolute distance) and with quadratic and linear priors for both the image and disparity. We used cliques belonging to the first order neighborhood in the prior terms for both image and disparity.

Figure 1 shows the results for the 'parking meter' stereo pair. This is the case of a predominantly slanting surface stereo imagery. For this example the L_1-norm was used as a data term. The prior terms $V_c^\delta(\delta)$ and $V_c^x(x)$ are linear functions, $V_c^\delta(\delta) = 1.5\,|\delta_{p,q} - \delta_{r,s}|$ and $V_c^x(x) = 0.1\,|x_{p,q} - x_{r,s}|$ where (p,q) and (r,s) are pixels belonging to the clique c. Starting with any arbitrary initialization for the HR disparity and the SR image, we minimize the cost function of (10) by SA, where the SR image label and the disparity label is updated iteratively for each point (p,q) on the HR grid. While updating the disparity, the SR image is fixed at its current estimate and vice-versa. The energy in (13) for each point (p,q) is updated locally, taking the advantage of Theorem 1. Fig 1(a) shows the low resolution observation. Fig 1(b) shows the bilinear interpolation output. Our SR image result is shown in Fig 1(c), whereas the HR disparity can be seen in Fig 1(d).

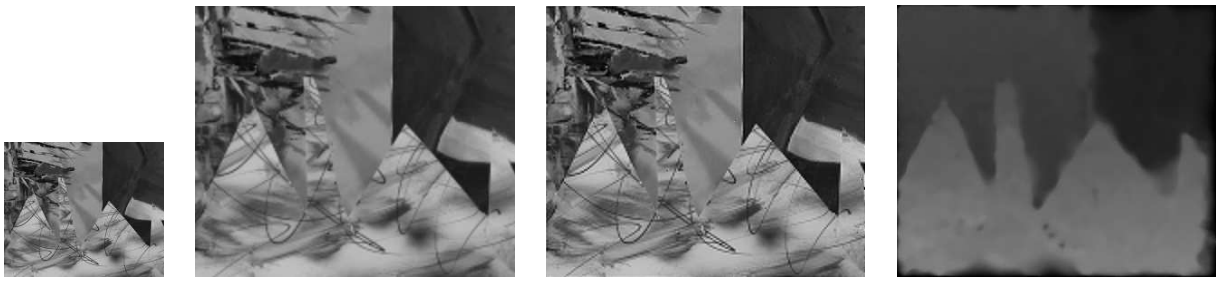
The fronto-parallel scene of the 'tsukuba' stereo pair is shown in Fig 2. Here, we choose absolute distance data term and linear priors both for image and disparity, except for the change in the weights of the priors. The prior terms used were $V_c^\delta(\delta) = |\delta_{p,q} - \delta_{r,s}|$ and $V_c^x(x) = 0.05\,|x_{p,q} - x_{r,s}|$. This scene contains more details than the previous one. Also, the surfaces are mainly fronto-parallel as opposed to the previous example. Hence, the weights on both the prior terms is reduced to avoid oversmoothing. One can clearly make out the difference between the interpolated image and SR result especially around the face region. Also at well textured regions, the HR disparity estimates are good. However, at textureless regions (such as the bottom of the desk), the disparity results tend to be oversmoothed by the nearby better textured regions.

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

Fig. 1. Parking meter stereo pair (a) LR observation (b) Bilinear interpolated image (c) Super-resolved image (d) HR disparity.

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

Fig. 2. Tsukuba stereo pair (a) LR observation (b) Bilinear interpolated image (c) Super-resolved image (d) HR disparity.



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

Fig. 3. Sawtooth stereo pair (a) LR observation (b) Bilinear interpolated image (c) Super-resolved image (d) HR disparity.

Fig 3 shows the 'sawtooth' stereo pair from the Middlebury stereo database [16]. In this experiment we used the squared distance data term and quadratic priors (also known as Gaussian MRF), $V_c^{\delta}(\delta) = (\delta_{p,q} - \delta_{r,s})^2$ and $V_c^{x}(x) = (x_{p,q} - x_{r,s})^2$. This is especially a good example for showing the super-resolution of the image. One can see that the textured details and the edges on the surfaces of the bottom sawtooth shapes have come out to be much sharper than in the interpolated version.

V. Conclusion

In this paper, we discussed a framework to simultaneously estimate the HR disparity and SR image from a set of LR stereo images. Our results clearly show the feasibility of such an approach. Although, our stress was on the idea of joint estimation rather than the solution methodology, we mention that a better solution approach, e.g. based on recent graph based methods could be explored. Moreover, physical aspects such as occlusion handling or estimation at slanting surfaces in such a integrated disparity estimation and super-resolution framework are some future prospects of this work. In summary, we look at this this work as a step towards taking super-resolution research in the 3d domain.

References

[1] K. V. Suresh and A. N. Rajagopalan, "Robust and computationally efficient super-resolution algorithm," *Journal of the Optical Society of America - A (JOSA-A)*, Vol. 24, No. 4, pp. 984-992, 2007.

[2] A. Rav-Acha, A. Zomet and S. Peleg, "Robust Super Resolution," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, Vol. 1, pp. 645-650, 2001.

[3] R. C. Hardie, K. J. Barnard and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp. 1621-1633, 1997.

[4] H. Shen, L. Zhang, B. Huang and P. Li, "A MAP approach for joint motion estimation, segmentation and superresolution," *IEEE Transactions on Image Processing*, Vol. 16, No. 2, pp. 479-490, 2007.

[5] S. Farsiu, D. Robinson, M. Elad and P. Milanfar, "Fast and robust super-resolution," *IEEE International Conference on Image Processing (ICIP 2003)*, Vol. 2, pp. 14-17, 2003.

[6] S. C. Park, M. K. Park and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, Vol. 20, No. 3, pp. 21-36, 2003.

[7] S. Chaudhuri and M. V. Joshi, “Motion free super-resolution,” *Springer-Verlag New York, Inc*, 2003.
[8] D. Rajan and S. Chaudhuri, “Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, pp. 1102-1117, 2003.
[9] A. N. Rajagopalan and V. P. Kiran, “Motion-free superresolution and the role of relative blur,” *Journal of the Optical Society of America - A (JOSA-A)*, Vol. 20, No. 11, pp. 2022-2032, 2003.
[10] M. V. Joshi and S. Chaudhuri, “Simultaneous estimation of super-resolved depth map and intensity field using photometric cue,”*Computer Vision and Image Understanding*, Vol. 101, No. 1, pp. 31-44, 2006.
[11] W. T. Freeman, T. R. Jones and E. C. Pasztor, “Example-based super-resolution” *IEEE Computer Graphics and Applications*, Vol. 22, No. 2, pp. 56-65, 2002.
[12] Z. Lin, J. He, X. Tang and C. Tang, “Limits of learning based super-resolution” *Technical Report MSR-TR-2007-92*, 2007.
[13] S. Baker and T. Kanade, “Super-resolution optical flow,” *Technical Report CMU-RI-TR-99-36*, 1999.
[14] R. Fransens, C. Strecha and L. Van Gool, “Optical flow based super-resolution: a probabilistic approach,” *Computer Vision and Image Understanding*, Vol. 106, No. 1, pp. 106-115, 2007.
[15] K. V. Suresh and A. N. Rajagopalan, “Robust space-variant super-resolution,” *Proc. IET International Conference on Visual Information Engineering (VIE 2006)*, pp. 600-605, 2006.
[16] vision.middlebury.edu/stereo/
[17] http://vasc.ri.cmu.edu/idb/html/jisct/index.html
[18] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, Vol. 47, No. 1/2/3, pp. 7-42, 2002.
[19] Y. Boykov, O. Veksler and R. Zabih, “Fast approximate energy minimizations via graph cuts,” *IEEE Transactions on Pattern Analysis and Machine Vision*, Vol. 23, No. 11, pp. 1222-1239, 2001.
[20] H. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts ?,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp. 147-159, 2007.
[21] A. Raj and R. Zabih, “A graph cut algorithm for generalized image deconvolution,” *IEEE International Conference on Computer Vision, (ICCV 2005)*, Vol. 2, pp. 1048-1054, 2005.
[22] U. Mudenagudi, R. Singla, P. Kalra and S. Banerjee, “Super resolution using graph-cut,” *Asian Conference on computer vision (ACCV 2006)*, Vol. 2, pp. 385-394, 2006.
[23] H. Hirschmller and D. Scharstein, “Evaluation of cost functions for stereo matching,” *In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, pp. 1-8, 2007.