Variational Multigrid for Fast 3D Interpretation of Image Sequences

Jong-Sung Kim and Ki-Sang Hong
Pohang University of Science and Technology
Image Information Processing Laboratory
San 31 Hyoja-Dong, Pohang, Korea
{kimjs,hongks}@postech.ac.kr

Abstract

We propose a variational multigrid method for fast 3D interpretation of image sequences, in which a dense depth map and 3D motion are directly recovered from spatio-temporal change of intensity images without prior matching and estimation. In this paper, we adopt the multigrid methods to efficiently reduce the computational complexity of the variational method, and suggest a new variational formulation to reliably perform the 3D interpretation. We show the efficiency and effectiveness of our method through experimental results with synthetic and real images.

1. Introduction

3D interpretation is an important problem in computer vision. Recovered 3D information can be used for several applications, such as robot navigation, surveillance, object recognition, video compression, and graphical rendering of visual data.

Most methods for 3D interpretation firstly compute sparse or dense correspondences in visual data sets [3, 4]. Such correspondences are performed by image features, stereo disparity, or optical flow. However, these approaches depend on the accuracy of matching or estimation process and are, in general, ill-posed. As an alternative, variational 3D interpretation was first studied in [5, 6], which only used the spatio-temporal change of intensity images to estimate scene depth and 3D motion and do not require prior matching and estimation. In these methods, 3D interpretation was stated as an energy minimization problem, and a partial differential equation (PDE) system was derived to obtain a solution of the problem. However, this approach leads to large-scale challenging numerical problems; e.g. there are more than two million variables in the PDE system to evaluate with a 640×480 video image [5]. Hence, variational 3D interpretation has been impractical for time-critical applications.

In this paper we adopt multigrid methods [2] for variational 3D interpretation to solve a PDE system with a fast convergence rate and linear computational complexity. Although multigrid methods are among the fastest methods for solving PDE systems, these schemes have not been applied in our area. However, multigrid methods do not provide a single algorithm for all problems. Hence, a new multigrid framework should be designed for solving our problem. For this end, we suggest a new variational formulation so that we can derive a linear PDE system with constant coefficients, which is more preferable to the nonlinear PDE system of the prior work from the viewpoint of the multigrid implementation. Coefficients are determined to preserve discontinuities in the multigrid solution of scene depth and motion. Then, we develop a multigrid scheme for efficiently solving the PDE system. We compare the efficiency and effectiveness of our method with the previous method.

The remainder of the paper is organized as follows. Section 2 introduces the basic concepts of variational 3D interpretation. Section 3 describes our variational formulation with its multigrid implementation. We close with experimental results and conclusions in Sections 4 and 5.

2. Problem Statement

Under the instantaneous motion model assumptions [5, 6], a 3D point $\mathbf{X}$ on a moving object satisfies the differential equation

$$\dot{\mathbf{X}} = \mathbf{v} + \mathbf{w} \times \mathbf{X}, \tag{1}$$

where $\mathbf{v} = (v_1, v_2, v_3)^\top$ and $\mathbf{w} = (w_1, w_2, w_3)^\top$ are the linear and angular velocities of the object motion, respectively. $\mathbf{X} = Z\mathbf{x}$ is used in (1), where Z is a depth value, and $\mathbf{x} = (x, y, 1)^\top$ a 2D image position. Through a pinhole camera, the image motion $\mathbf{u}$, corresponding to $\mathbf{X}$, is

$$\mathbf{u} = Z^{-1}C\mathbf{v} + D\mathbf{w} \equiv C\mathbf{t} + D\mathbf{w}, \tag{2}$$

where $\mathbf{t} = (t_1, t_2, t_3)^\top = Z^{-1}\mathbf{v}$, and the definitions of C and D can be found in [5]. As we can see in (2), scene depth

can be recovered only up to the scale factor [5, 6]. Hence, we can only determine the inverse value of scene depth as $Z^{-1} = ||\mathbf{t}||$ by assuming that $||\mathbf{v}|| = 1$.

Let I_0 and I_1 be two consecutive frames of a image sequence. Then, by using the constant brightness assumption and (2), we can obtain an expression that relates the 3D variables to the image spatio-temporal derivaties:

$$I_1 + \mathbf{c}^\top \mathbf{t} + \mathbf{d}^\top \mathbf{w} - I_0 = 0, \tag{3}$$

where $\mathbf{c} = \nabla I_1^\top C$ and $\mathbf{d} = \nabla I_1^\top D$. Using this fact, H. Sekkati et al. [6] developed a 3D interpretation that minimizes the following energy functional:

$$E_1(\mathbf{t}, \mathbf{w}) = \int\int_\Omega \big(I_1 + \mathbf{c}^\top \mathbf{t} + \mathbf{d}^\top \mathbf{w} - I_0\big)^2 dxdy \tag{4}$$
$$+\lambda \int\int_\Omega \sum_{i=1}^{3} \Big[\Phi\big(||\nabla t_i||\big) + \Phi\big(||\nabla w_i||\big)\Big] dxdy,$$

where $\Phi(s) = 2\sqrt{1+s^2} - 2$. The Euler-Lagrange equations for t_i, w_i, $i = 1, 2, 3$ are given by

$$\begin{cases} \lambda \mathrm{div}\big(\frac{\Phi'\left(||\nabla t_i||\right)}{||\nabla t_i||}\nabla t_i\big) = 2c_i\big(I_1 + \mathbf{c}^\top \mathbf{t} + \mathbf{d}^\top \mathbf{w} - I_0\big) \\ \lambda \mathrm{div}\big(\frac{\Phi'\left(||\nabla w_i||\right)}{||\nabla w_i||}\nabla w_i\big) = 2d_i\big(I_1 + \mathbf{c}^\top \mathbf{t} + \mathbf{d}^\top \mathbf{w} - I_0\big) \end{cases} \tag{5}$$

with boundary conditions $\partial_{\mathbf{n}} t_i = 0$, $\partial_{\mathbf{n}} w_i = 0$, $i = 1, 2, 3$, where $\mathbf{n}$ is the unit normal to the boundary $\partial\Omega$ of Ω. A discretization of (5) yields a large-scale system of non-linear PDEs which is difficult to solve. Gauss-Seidel relaxation (GSR) iterations were used within continuation as in a half-quadratic (HQ) minimization algorithm [1], which should alternatively compute auxiliary variables. However, the computational cost of this procedure is very high because of the slow convergence rate of GSR and the additional cost for updating the auxiliary variables.

3. Variational Multigrid for 3D interpretation

We develop a new variational formulation and fast numerical implementations based on multigrid methods, which requires on the order of $O(N)$ operations, compared to typical complexities of $O(N^3)$ operations of a basic relaxation, where N is the number of nodes in the current grid.

3.1. Variational Formulation

We minimize an energy functional $E_2 = M + R$, where $M \equiv M(\mathbf{t}, \mathbf{w})$ measures the conformity to data according to (3), and $R \equiv R(\mathbf{t}, \mathbf{w})$ defines regularizing constraints on $\mathbf{t}$ and $\mathbf{w}$. Considering the measurement noise and the approximation error in (3), we define a spatially smoothed gradient constraint equation

$$I_1^\sigma + \mathbf{c}^{\sigma\top} \mathbf{t} + \mathbf{d}^{\sigma\top} \mathbf{w} - I_0^\sigma = \delta, \tag{6}$$

where $I_0^\sigma = I_0 * G_\sigma$, $I_1^\sigma = I_1 * G_\sigma$, $\mathbf{c}^\sigma = \nabla I_1^{\sigma\top} C$, $\mathbf{d}^\sigma = \nabla I_1^{\sigma\top} D$, δ is a error variable, and $I_\sigma = G_\sigma * I_j$ represents the convolution of I_j with a Gaussian of standard deviation σ. Hence, the data term M in our energy functional is

$$M(\mathbf{t}, \mathbf{w}) = \frac{1}{2\sigma_\delta^2} \int\int_\Omega \big(I_1^\sigma + \mathbf{c}^{\sigma\top} \mathbf{t} + \mathbf{d}^{\sigma\top} \mathbf{w} - I_0^\sigma\big)^2 dxdy. \tag{7}$$

The regularization term R is defined so that it realizes a data-driven anisotropic regularization [7] to preserve the discontinuities in scene depth and motion and ease implementation with multigrid methods. Thus, we introduce a variant of diffusion tensor

$$T\big(\nabla I_0^\sigma\big) = \frac{1}{||\nabla I_0^\sigma||^2 + 2\nu^2} \begin{bmatrix} |\partial_y I_0^\sigma|^2 + \nu^2 & 0 \\ 0 & |\partial_x I_0^\sigma|^2 + \nu^2 \end{bmatrix}, \tag{8}$$

where ν is a positive constant, and then we define a first-order regularization R_1 by

$$R_1(\mathbf{t}, \mathbf{w}) = \int\int_\Omega \sum_{i=1}^{3} \Big[\nabla t_i^\top T(\nabla I_0^\sigma) \nabla t_i + \nabla w_i^\top T(\nabla I_0^\sigma) \nabla w_i\Big] dxdy. \tag{9}$$

We also propose a zero-order regularization R_0, defined as

$$R_0(\mathbf{t}, \mathbf{w}) = \int\int_\Omega \frac{1}{||\nabla I_0^\sigma||^2 + 2\nu^2} \sum_{i=1}^{3} \big(t_i^2 + w_i^2\big) dxdy, \tag{10}$$

which is introduced to suppress the value of 3D variables where texture is not available. (7), (10), and (9) lead to the following energy functional:

$$E_2(\mathbf{t}, \mathbf{w}) = M(\mathbf{t}, \mathbf{w}) + \mu R_0(\mathbf{t}, \mathbf{w}) + \lambda R_1(\mathbf{t}, \mathbf{w}). \tag{11}$$

The Euler-Lagrange equations for t_i, w_i, $i = 1, 2, 3$, corresponding to (11), are given by

$$\begin{cases} \lambda \mathrm{div}\big(T(\nabla I_0^\sigma)\nabla t_i\big) = \\ \quad \frac{c_i^\sigma}{\sigma_\delta^2}\big(I_1^\sigma + \mathbf{c}^{\sigma\top} \mathbf{t} + \mathbf{d}^{\sigma\top} \mathbf{w} - I_0^\sigma\big) + \frac{2\mu t_i}{||\nabla I_0^\sigma||^2 + 2\nu^2} \\ \lambda \mathrm{div}\big(T(\nabla I_0^\sigma)\nabla w_i\big) = \\ \quad \frac{d_i^\sigma}{\sigma_\delta^2}\big(I_1^\sigma + \mathbf{c}^{\sigma\top} \mathbf{t} + \mathbf{d}^{\sigma\top} \mathbf{w} - I_0^\sigma\big) + \frac{2\mu w_i}{||\nabla I_0^\sigma||^2 + 2\nu^2} \end{cases} \tag{12}$$

We obtain a linear PDE system by discretizing the Euler-Lagrange equations in (12) via the finite difference approximation [2]. Then, we apply the multigrid methods to the linear PDE system. The discrete Euler-Lagrange equations corresponding to (12) are not included to save space.

3.2. Full Multigrid Implementation

Let

$$A_h z_h = b_h \tag{13}$$

describe a linear PDE system for 3D interpretation, where A_h is the stiffness matrix, b_h the source term, the subscript $_h$ the size of the grid spacing, and z the exact solution. Since $\mathbf{t}$ and $\mathbf{w}$ are independent of each other, we obtain linear systems for each. Therefore, z_h is a $3N$-dimensional vector consisting of $\mathbf{t}$ and $\mathbf{w}$ of all image pixels in lexicographical order, and A_h is a $3N \times 3N$ matrix, of which each element comes from the discrete Euler-Lagrange equations. We define $\tilde{z}$ with a current or next approximate solution, and $\bar{z}$ with a smoothed or corrected solution. Combining the ideas of relaxation and coarse-grid correction, we can obtain the two-grid iteration:

1. *Pre-smoothing*: Compute $\bar{z}_h$ by applying μ_1 sweeps of a relaxation method to $\tilde{z}_h$.
2. *Coarse-grid correction*:
   (a) Compute the residual $r_h = b_h - A_h \bar{z}_h$ on the fine grid.
   (b) Restrict the residual $r_{2h} = R_h^{2h} r_h$ to the coarse grid.
   (c) Solve exactly the system $A_{2h} e_{2h} = r_{2h}$ on the coarse grid for the correction.
   (d) Prolong the correction to the fine grid by $e_h = P_{2h}^h e_{2h}$.
   (e) Correct the next approximation by $\bar{z}_h \leftarrow \bar{z}_h + e_h$.
3. *Post-smoothing*: Compute $\tilde{z}_h$ by applying μ_2 sweeps of a relaxation method to $\bar{z}_h$.

We can apply this two-grid method recursively down to some coarsest grid, denoted by the subscript $_H$, where the solution can be found easily by iterating the relaxation scheme to convergence. This recursive scheme is the multigrid method. One iteration of a multigrid method is called a multigrid V-cycle (MV). Efficiency can be improved by using the full multigrid algorithm (FMG). Instead of starting with $\tilde{z}_h = 0$, the first approximation is interpolated from a coarse-grid solution, which is found by a similar FMG process from even coarser grids. The FMG for the V-cycle (FMV) is defined recursively by the following two steps:

1. If $h' = H$, then solve exactly by $\bar{z}_H = A_H^{-1} b_H$. Otherwise, apply a FMV to $\bar{z}_{2h'}$.
2. Set initial guess by $\bar{z}_{h'} = P_{2h'}^{h'} \bar{z}_{2h'}$, and compute $\tilde{z}_{h'}$ by applying η sweeps of a MV to $\bar{z}_{h'}$.

4. Experimental Results

Numerical experiments were conducted with synthetic and real image sequences. All computations were carried out on a desktop PC with a 2.4-GHz Intel P4 CPU executing C/C++ code. We evaluated the recovered 3D information by using a display of gray value rendering of the scene depth map as in [5, 6]. Since scene depth can be recovered up to scale, we used the scaled inverse value of scene depth, sZ^{-1} , for rendering a gray level map,

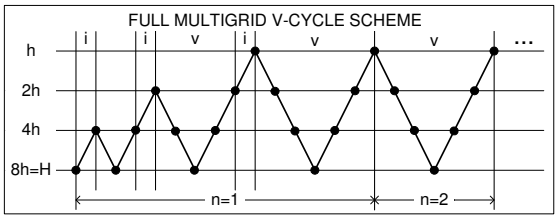

Figure 1. Our FMV scheme. V-cycles are denoted with v, interpolation steps with i, and iteration numbers with n.

Table 1. Comparison of the numerical performance for variational 3D interpretation.

Scheme	n	r_{rel}	Runtime [s]
GSR,HQ,E_1[6]	506	0.000998	136.88
	1299	0.000009	349.44
GSR,E_2	304	0.000998	53.85
	559	0.000009	100.09
FMV,HQ,E_1	4	0.000633	11.73
	11	0.000007	26.45
FMV,E_2	5	0.000358	13.53
	9	0.000004	21.33

where $s = 255/\max(\epsilon, ||\mathbf{t}||)$ was used with a small positive value ϵ. For the numerical evaluation, we compared the L_2-norm of residual $||r_h||_2$ and the relative residual $r_{rel} = ||r_h^n||_2 / ||r_h^0||_2$ of each scheme. σ_δ in (7) was estimated in a robust and simple way as $\sigma_\delta = \mathrm{median}(|\delta|)/\sqrt{2}$. The FMV scheme in Figure 1 was used in all experiments with four grids, $\mu_1 = 1$, $\mu_2 = 1$, and $\eta = 1$.

We evaluated the numerical performance of our method by synthesizing a 320×240 image pair as in [6]. Both E_1 and E_2 were tested for variational 3D interpretation. Table 1 shows the required number of iterations n to reach the desired precisions of $r_{rel} < 10^{-3}$ and 10^{-5}, the actual value of the relative residual, and the runtime for each scheme. Figure 2 clearly shows that our FMV scheme converges faster than the GSR scheme.

Figure 3 compared the effectiveness of the proposed variational formulation E_2 to E_1 within the multigrid framework. Using another synthetic image pair of a dynamic scene of three independently moving squares. E_2 preserved the discontinuities in depth better that E_1 do when the full multigrid scheme was adopted.

Our method was applied to two test examples, the *Rotating cube* sequence (256×240) and the *Yosemite* sequence (316×252). Figures 4 and 5 show depth maps obtained by

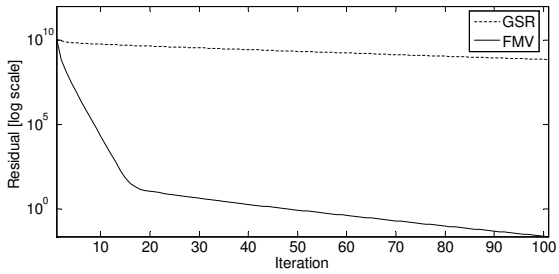

Figure 2. Comparison of (GSR,$E2$) and (FMV,$E2$) in terms of log-scaled $||r_h^n||_2$.

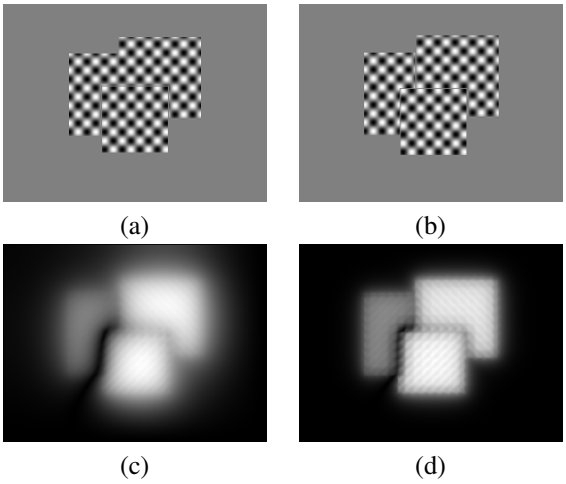

Figure 3. Simulation results. (a), (b) Synthetic image pair. (c), (d) Depth maps obtained by five FMV iterations with E_1 and E_2, respectively.

our method and a 3D rendering result of textured depth map to visually demonstrate the real performance of our method. In Figures 4(c) and 4(d), the total computations took 17.9s and 24.05s, respectively.

5. Summary and Conclusions

We proposed a variational multigrid method for fast 3D interpretation of image sequences. The previous method has been very time-consuming, because it could not efficiently deal with the large scale problem of the variational computations. We applied multigrid methods and suggested a new variational formulation to efficiently solve the 3D interpretation problem. We have experimentally verified the superiority of our method.

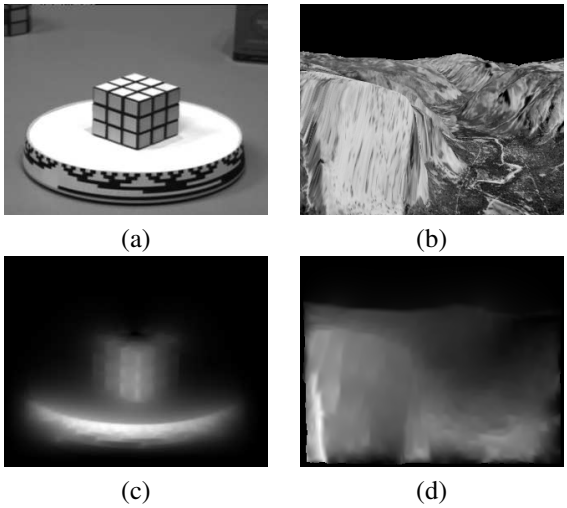

Figure 4. Test examples. (a) *Rotating cube* **sequence. (b)** *Yosemite* **sequence. (c), (d) Depth maps obtained by ten FMV iterations with E_2.**

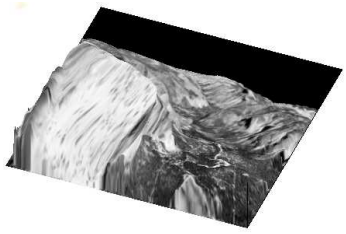

Figure 5. A 3D rendering result of the textured depth map of the *Yosemite* **sequence.**

References

[1] G. Aubert and P. Kornprobst. *Mathematical problems in image processing*. Springer Verlag, 2002.

[2] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, 2nd edition, 2000.

[3] R. Chellappa, G. Qian, and S. Srinivasan. Structure from motion: sparse versus dense correspondence methods. In *Proc. ICIP*, volume 2, pages 492–499, 1999.

[4] T. S. Huang and A. N. Netravali. Motion and structure from feature correspondences: A review. In *Proceedings of the IEEE*, volume 82(2), pages 252–268, 1994.

[5] A. Mitiche and S. Hadjres. Mdl estimation of a dense map of relative depth and 3d motion from a temporal sequence of images. *Pattern Analysis and Applications*, 6:78–87, 2003.

[6] H. Sekkati and A. Mitiche. Dense 3d interpretation of image sequences: A variational approach using anisotropic diffusion. In *Proc. ICIAP*, 2003.

[7] J. Weickert and H. Scharr. A theoretical framework for convex regularizers in pde-based computation of image motion. *Int'l J. Computer Vision*, 45(3):103–118, 2001.