NAIST-IS-TR94027
ISSN 0919-9527

Dense Matching of Two Views with Large Displacement[1]

Naokazu Yokoya[2]

August 1994

ABSTRACT

This paper describes a method of establishing dense matching of two views with large displacements. The problem addressed is formulated as the minimization of an energy functional that combines a similarity term and a smoothness term. The minimization of the energy functional reduces to solving a large system of nonlinear equations which is a discretized version of the Euler-Lagrange equation for the energy minimization problem at each image point. A dense displacement map is computed by solving the system of equations using a coarse-to-fine approach. The method has been successfully experimented with two applications: (1) computing binocular disparities from stereo pair images and (2) computing a dense displacement vector field (optical flow) from two views in a time-varying image sequence.

Keywords: computer vision, matching, regularization, energy minimization, binocular stereo, optical flow

[1]This work was supported in part by the Real World Computing Program of MITI under a subcontract from the Real World Computing Partnership and also by the Tateishi Science and Technology Foundation.

[2]Information Technology Center, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-01, Japan
E-mail: yokoya@itc.aist-nara.ac.jp

1 Introduction

Matching of two views is one of the most important problems in computer vision; especially in binocular stereo and motion analysis [7]. This problem requires to establish correspondences between images. Most existing techniques for finding correspondences between images can be classified into two categories: gradient method and feature-based matching method.

The gradient method is based on the assumptions that the intensity of a point is kept unchanged from image to image and the displacement of each point is sufficiently small between images. From such assumptions, the displacement (u, v) at each point (x, y) in the image $I(x, y, t)$ is approximately constrained by the following equation [2].

$$\frac{\partial I(x,y,t)}{\partial x}u + \frac{\partial I(x,y,t)}{\partial y}v + \frac{\partial I(x,y,t)}{\partial t} = 0, \tag{1}$$

where two different views are represented by time axis t ($I(x, y, 1)$ and $I(x, y, 2)$). The displacement (u, v) is uniquely determined by imposing an additional smoothness constraint on the displacement (u, v); that is, the problem of computing (u, v) is formulated in the framework of *regularization* [4]. Numerical computation is usually carried out by using a relaxation technique. This approach can compute a dense displacement field (optical flow) without feature extraction and explicit matching procedure. However, it should be noted that the differentiability of the image with respect to time axis as well as image coordinates is required.

The feature-based matching method consists of feature extraction and explicit feature matching, sometimes using geometric constraints [1, 3]. There are a variety of features; interest points with high intensity variation, edge elements with large magnitude of intensity gradient, zero-crossings of a Laplacian-Gaussian filtered image, and line segments as connected edge elements. In this type of approach, the displacements are computed sparsely because correspondences between images are established only at features. In contrast with the gradient method, the feature-based method has the advantage that it can compute large displacements between images.

This paper describes a new gradient method which can compute a dense displacement vector field from two views with large displacements. The problem addressed is formulated as the minimization of an energy functional that combines a similarity term for corresponding points and a smoothness term for a solution. The minimization of the energy functional reduces to solving a large system of nonlinear equations which is a discretized version of the Euler-Lagrange equation for the energy minimization problem at each image point. A dense displacement map is computed by solving the system of equations using a coarse-to-fine approach. From an initial estimate of displacement field, the proposed algorithm first finds a local minimum of the energy functional at a coarse scale and then proceeds to

finer scales by using the previous result as an initial estimate at the current scale. In the following, the approach is described in detail and experimental results are presented for two applications: (1) computing binocular disparities from stereo pair images and (2) computing a dense displacement field from two views in a time-varying image sequence.

2 Formulation of Matching of Two Views

Two views are denoted by $I_1(x, y)$ and $I_2(x, y)$, respectively. The displacement is referred to as $(u(x, y), v(x, y))$ at a point (x, y) in the first view; that is, the point (x, y) in the first view is assumed to correspond to the point $(x + u(x, y), y + v(x, y))$ in the second view (See Fig. 1).

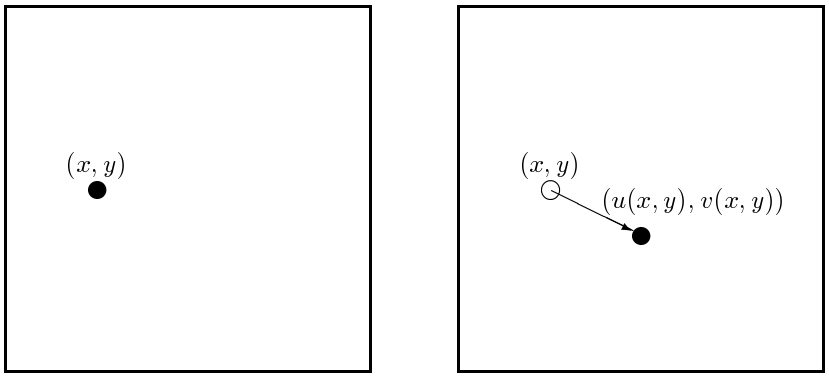


Figure 1: Displacement between two views.

The problem here is to determine the displacement $(u(x, y), v(x, y))$ which is assumed to be a vector of real-valued. By using the framework of regularization, the problem is formulated in a continuous form as that of finding a pair of functions ($u(x, y)$ and $v(x, y)$) which minimize the following energy functional:

$$E(u, v) = P(u, v) + \lambda S(u, v), \tag{2}$$

where $P(u, v)$ is a *penalty functional* which evaluates the difference in image feature values between corresponding points, and $S(u, v)$ is a *stabilizing functional* by which a *smoothness constraint* is imposed upon the solution u and v. The *regularization parameter* λ controls the strength of the smoothness constraint. We define the penalty and stabilizing functionals for the dense two-view matching problem as follows:

$$P(u, v) = \int\int \{I_1(x, y) - I_2(x + u(x, y), y + v(x, y))\}^2 \, dxdy, \tag{3}$$

$$S(u, v) = \int\int \left\{ \left(\frac{\partial u(x, y)}{\partial x} \right)^2 + \left(\frac{\partial u(x, y)}{\partial y} \right)^2 + \left(\frac{\partial v(x, y)}{\partial x} \right)^2 + \left(\frac{\partial v(x, y)}{\partial y} \right)^2 \right\} dxdy. \tag{4}$$

It should be noted that the membrane spline is employed to embed the smoothness constraint on each component of displacement vector.

3 Numerical Solution

Now we describe how to solve the energy minimization problem defined in the previous section. The *calculus of variations* tells us that the necessary condition of the minimization of Eq. (2) is a couple of the following Euler-Lagrange equations:

$$\Big\{I_2(x+u(x,y),y+v(x,y))-I_1(x,y)\Big\}\cdot\frac{\partial I_2(x+u(x,y),y+v(x,y))}{\partial x} -\lambda\left\{\frac{\partial^2 u(x,y)}{\partial x^2}+\frac{\partial^2 u(x,y)}{\partial y^2}\right\}=0, \tag{5}$$

$$\Big\{I_2(x+u(x,y),y+v(x,y))-I_1(x,y)\Big\}\cdot\frac{\partial I_2(x+u(x,y),y+v(x,y))}{\partial y} -\lambda\left\{\frac{\partial^2 v(x,y)}{\partial x^2}+\frac{\partial^2 v(x,y)}{\partial y^2}\right\}=0. \tag{6}$$

By discretizing Eqs. (5) and (6) at each grid (pixel) (i,j) using finite difference approximations, we obtain a Gauss-Seidel type relaxation:

$$u^{(k)}(i,j)=\frac{1}{4}\Big[\{u^{(k-1)}(i+1,j)+u^{(k)}(i-1,j)+u^{(k-1)}(i,j+1)+u^{(k)}(i,j-1)\} -\frac{t^2}{\lambda}\{I_2(i+u^{(k-1)}(i,j),j+v^{(k-1)}(i,j))-I_1(i,j)\}\frac{\partial I_2(i+u^{(k-1)}(i,j),j+v^{(k-1)}(i,j))}{\partial x}\Big], \tag{7}$$

$$v^{(k)}(i,j)=\frac{1}{4}\Big[\{v^{(k-1)}(i+1,j)+v^{(k)}(i-1,j)+v^{(k-1)}(i,j+1)+v^{(k)}(i,j-1)\} -\frac{t^2}{\lambda}\{I_2(i+u^{(k-1)}(i,j),j+v^{(k-1)}(i,j))-I_1(i,j)\}\cdot\frac{\partial I_2(i+u^{(k-1)}(i,j),j+v^{(k-1)}(i,j))}{\partial y}\Big], \tag{8}$$

where t represents the distance between two adjacent pixels and the superscript k denotes the iteration step.

Since $u(i,j)$ and $v(i,j)$ are assumed to be real-valued, a point $(i+u(i,j),j+v(i,j))$ in the second image does not necessarily lie at an integer grid, and hence we have to calculate intensity values at subpixel locations. We estimate image intensity values at subpixel locations by locally fitting bi-cubic B-spline surfaces to image intensities at pixels.

4 Iterative Computation of Displacements over Scale

By using Eqs. (7) and (8), we can iteratively estimate the displacement $(u(i,j),v(i,j))$ from an initial estimate. However we encounter a difficulty in this iterative computaion when we

cannot select a good initial estimate; that is, we cannot necessarily obtain a global minimum of the energy in Eq. (2) from an arbitrarily selected initial estimate, because Eq. (2) generally has a lot of local minima. In order to avoid local minima, We employ a multiscale algorithm [8] using pyramidal images. The pyramid consists of $L+1$ layers of dimension $2^{L-l} \times 2^{L-l}$ stacked one upon another, with the finest image of dimension $2^L \times 2^L$ at the bottom (level 0) and the $2^0 \times 2^0$ single point image at the apex (level L). Each pixel at level l has its four predecessors at level $l-1$. In our implementation, the level l of the pyramid is the Gaussian-filtered version of the original image with the characteristic size $\sigma = 2^l$.

Our multiscale algorithm using Eqs. (7) and (8) goes down through the pyramids from a starting (coarsest) layer to the bottom (finest) layer, by using the previous result as an initial estimate at the next layer. The algorithm is described in more detail in the following steps:

(1) [Initialization]

First, we select a starting layer l_0 of the pyramids, and then set the initial values of displacement functions: $u(i,j) = constant$ and $v(i,j) = constant$ for all the pixels.

(2) [Multiscale Estimation]

(2-a) At the start of each scale, the displacement functions obtained at the previous scale are projected down to the current scale.

(2-b) Displacement components $u(i,j)$ and $v(i,j)$ are then iteratively updated using Eqs. (7) and (8) until convergence.

(2-c) The computation at (2-a) and (2-b) is iterated until the finest scale is reached.

The starting layer l_0 should be selected such that $2^{l_0} \geq$ maximum displacement when the zero displacement is used as an initial estimate at level l_0. The distance t between two adjacent pixels included in Eqs. (7) and (8) is set to 2^l at level l of the pyramid.

Note that such a multiscale technique as described above can significantly improve the efficiency of the relaxation method, in addition to avoiding meaningless local minima.

5 Experiments

We show the experimental results for two applications; one is to compute disparities from binocular stereo images and the other is to estimate a dense displacement vector field from two views in a time-varying image sequence. In experiments shown below, level 4 of the pyramid was selected as the starting layer (coarsest layer) of the algorithm. The termination of the iterative computation at each scale was decided according to the decreasing rate of

total energy changes in consecutive iteration steps as follows:

$$\frac{E_{k-1} - E_k}{E_k} < 0.0001. \tag{9}$$

The regularization parameter λ was set to 1000 and the zero displacement is used as an initial estimate at each pixel at level l_0 in all the experiments given below.

5.1 Computing Binocular Stereo Disparities

First the proposed method is applied to binocular stereo images for computing disparities, in which a modified version of the algorithm is used. In this application, the displacement considered is limited to horizontal disparity $u(x, y)$ because we employ a standard stereo camera geometry where the epipolar line coincides with the scanline and thus corresponding points between two images necessarily lie on the same scanline.

We define the penalty and stabilizing functionals for stereo matching as follows [8]:

$$P(u) = \int\int \alpha(x, y) \{I_1(x, y) - I_2(x + u(x, y), y)\}^2 \, dxdy, \tag{10}$$

$$S(u) = \int\int \beta(x, y) \left\{ \left(\frac{\partial u(x, y)}{\partial x} \right)^2 + \left(\frac{\partial u(x, y)}{\partial y} \right)^2 \right\} dxdy. \tag{11}$$

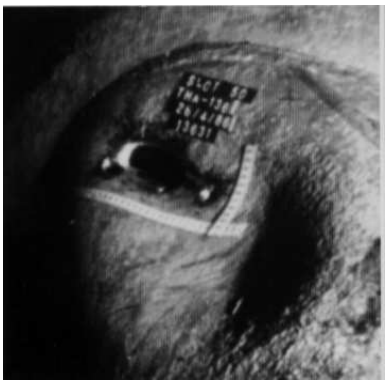
Note that we use two binary functions: the visibility function $\alpha(x, y)$ [8] and the continuity control function $\beta(x, y)$ [6]. This definition of penalty and stabilizing functionals enables us to consider occlusions and preserve disparity discontinuities. The continuity control function and visibility function are estimated in the process of multiscale estimation of disparities (See [8] for detail).

Figure 2 shows an experimental result with a stereo pair of a damaged underwater oil platform. It can be seen in Fig. 2 that the algorithm has excellently determined dense matching of binocular stereo images and thereby obtained a disparity surface of damaged concave structure.

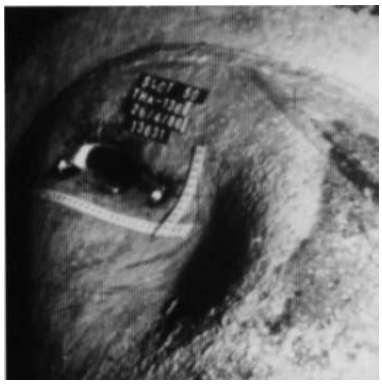
5.2 Computing Optical Flow from Two Views

The second experiment is to compute both horizontal and vertical disparities from a pair of images; that is, the problem is to compute a 2D displacement vector field (optical flow). The algorithm used in this experiment is exactly same as described in Section 3.

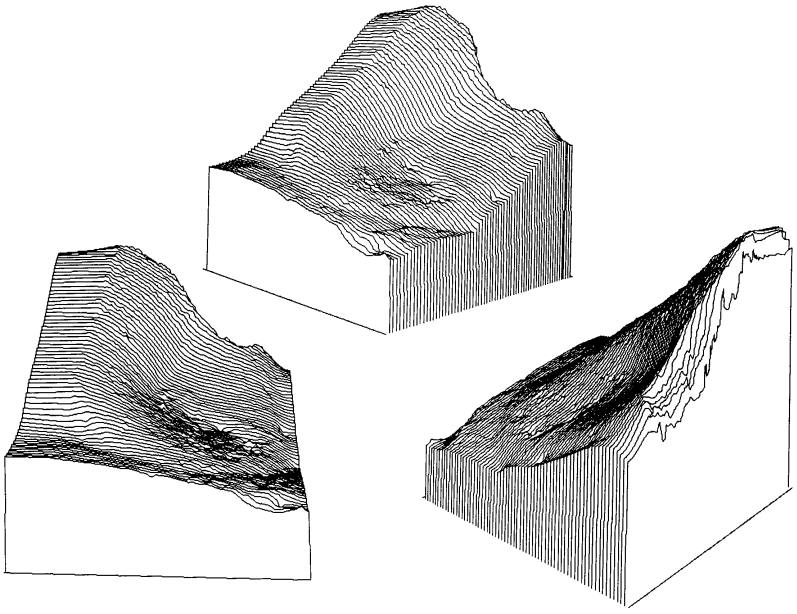
Figure 3 shows results of computing displacements from two-view images of a find place of historic interest. Computed disparities in both horizontal and vertical directions are coded in intensity and are also illustrated as 3D reliefs in Fig. 3. Two input images in this experiment look like a pair of binocular stereo images obtained by a standard stereo camera geometry



Left (second) image



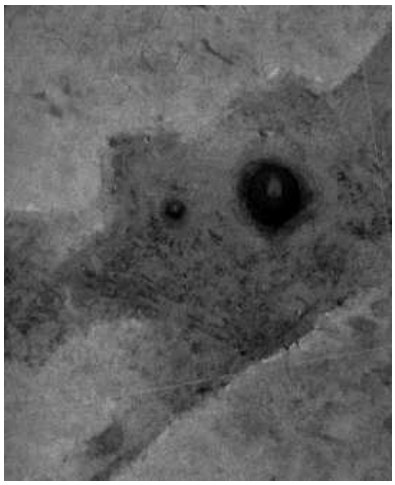
Right (first) image



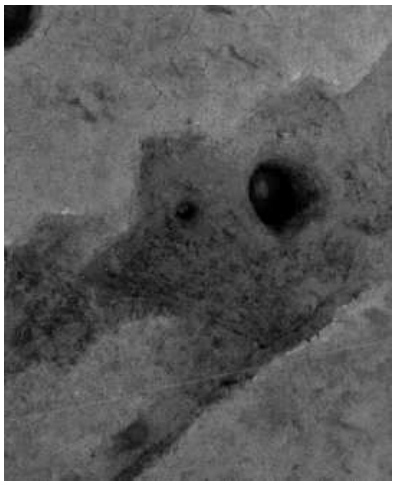
3D relief of computed disparities viewed from three different positions

Right image generated from left image using disparities

Figure 2: Computing disparities from binocular stereo images.



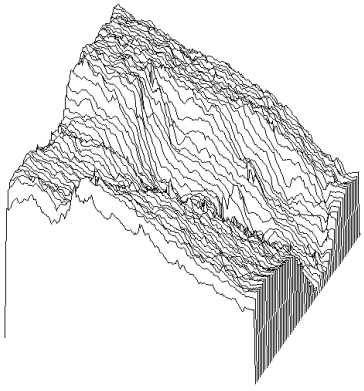
First image



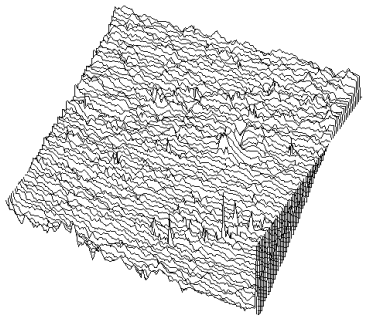
Second image

Computed horizontal disparities

Computed vertical disparities



3D relief of computed horizontal disparities

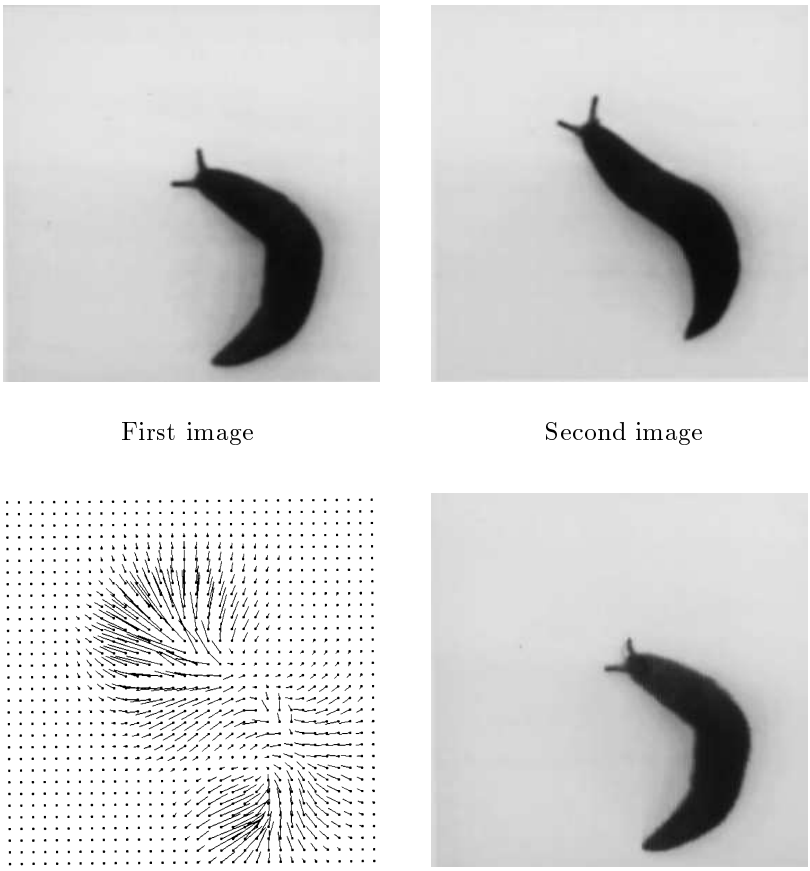


3D relief of computed vertical disparities

Figure 3: Computing displacements from two-view images obtained by camera motion.

with parallel viewing axes, however, the algorithm has clearly shown that the images contain vertical disparities caused by nontranslational camera motion.

Another example of determining a 2D displacement vector field is shown in Fig. 4. The input images are taken among a time-varying image sequence used in biological study [5], in which a deformable object (*slug*) is in motion.



First image

Second image

Computed displacement vector field (coarse scale)

First image generated from second image using displacements

Figure 4: Computing optical flow from two time-varying images of a moving deformable object.

The experiments including the results given above have indicated that the proposed method can correctly find correspondences between two views with large displacements (disparities).

6 Conclusions

We have proposed a new method of matching two views. The method has been successfully experimented with two typical applications: (1) computing binocular stereo disparities and (2) determining optical flow from two-view images. The problem of dense matching of two

views is formulated as that of energy minimization based on the framework of regularization. A coarse-to-fine approach using Gaussian pyramids is employed for solving the energy minimization problem.

The advantages of the proposed method can be summarized as follows:

(1) Our method can obtain dense matching of two images; that is, it can determine correspondence at each pixel.

(2) Displacement at each pixel is estimated in subpixel accuracy by computing a displacement vector of real-valued.

(3) The method works well for images with large displacements as well as those with small displacements.

Acknowledgements

The author thanks Haruo Matsuo and Yoh Tamura of Nara Institute of Science and Technology for their assistance in experiments. We are most grateful to Richard Baldwin of the Polytechnic of East London and Sunao Katada of the Tezukayama University for providing us with stereo image data.

References

[1] S.T.Barnard and W.B. Thompson: "Disparity analysis of images", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.PAMI-2, No.4, pp.333-340, 1980.

[2] B.K.P. Horn and B.G. Schunck: "Determining optical flow", *Artificial Intelligence*, Vol.17, pp.185-203, 1981.

[3] Y. Ohta and T. Kanade: "Stereo by intra- and inter-scanline search using dynamic programming", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.PAMI-7, No.2, pp.139-154, 1985.

[4] T. Poggio, V. Torre, and C. Koch: "Computational vision and regularization theory", *Nature*, Vol.317, No.6035, pp.314-319, 1985.

[5] T. Sekiguchi, A. Yamada, H. Suzuki, and A. Mizukami: "Temporal analysis of the retention of a food-aversive conditioning in *Limax flavus*", *Zool. Sci.*, Vol.8, pp.103-111, 1991.

[6] D. Terzopoulos: "Regularization of inverse visual problems involving discontinuities", *IEEE Trans. Pattern Anal. Mach Intell.*, Vol.PAMI-8, No.4, pp.413-424, July 1986.

[7] J. Weng, N. Ahuja, and T.S. Huang: "Matching two perspective views", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.14, No.8, pp.806-825, 1992.

[8] N. Yokoya: "Surface reconstruction directly from binocular stereo images by multiscale-multistage regularization", *Proc. 11th Int. Conf. on Pattern Recognition*, Vol.I, pp.642-646, 1992.