Stereo Matching with Direct Surface Orientation Recovery

Hiroshi Hattori and Atsuto Maki
TOSHIBA Kansai Research Laboratories
8-6-26 Motoyama-Minami-Cho, Higashinada-ku, Kobe 6580015, Japan
[kan|maki]@krl.toshiba.co.jp

Abstract

We propose a new stereo matching algorithm which computes the depth and surface orientation simultaneously. While area-based stereo matching is an essential technique in the recovery of the dense 3D shape, conventionally it uses square windows based on the implicit assumption that intensity patterns surrounding corresponding points have no deformations between images from different views. In practice, however, the local surface orientations deform the intensity patterns and such deformations often give rise to poor estimation of the 3D shape. To solve this problem, we formulate a new algorithm that allows a matching window to locally deform according to the surface orientation, which we propose to compute directly from intensity gradients within the window. Through experiments we demonstrate that our algorithm indeed realizes more precise recovery of the 3D shape than do conventional ones while being applicable to various images.

1 Introduction

Recovering a 3D shape of an object or a scene has been a central issue in computer vision. In order to extract 3D information from 2D images, it is effective to use multiple images captured from different viewpoints. The essential problem is to find corresponding points between different images. When the viewing geometry is already known, i.e.,the geometric relationship between the images including the camera position and orientation, we can robustly recover the 3D shape since it is then attributed to a 1D search problem by exploiting the geometric relationship. This is called the epipolar constraint and the image matching technique using the epipolar constraint is often called stereo matching.

In recent decades numerous stereo matching algorithms have been proposed[18]. They can roughly be classified into two categories; feature-based and area-based. In the feature-based ones, features such as edges, lines or corners are extracted and matched. Although some robust depth estimation is often acquired using these approaches, the output depth information is inherently rather sparse. On the other hand, the characteristic of the area-based approaches is to recover the dense 3D shape by correlating the grey levels within the window around each pixel between different images. In this paper, for the purpose of dense 3D shape recovery, we consider the area-based approaches.

Conventional area-based stereo matching employs a “square window” to measure the similarity or non-similarity among the different image regions. This is based on the implicit assumption that the intensity patterns surrounding the corresponding points have

BMVC 1998 doi:10.5244/C.12.36

no deformations between images from different views. In general, however, as the local intensity patterns deform according to both the viewing geometry and the local surface orientations(or the depth gradients), the resulting deformations are not negligible in the stereo matching. Whereas the former deformation parameters are determined globally for the entire image, the latter parameters vary with image region. Therefore, a window must be locally deformed depending on the surface orientation. In other words, the conventional approaches have been associated with recovery of depth only, assuming that the 3D surface is locally fronto-parallel, that is, the local 3D surface and the image plane are parallel to each other. Although this assumption is valid if the window size is small enough, too small a window does not cover sufficient intensity variation to establish the correspondence, resulting in poor depth estimation.

This paper presents a new stereo matching algorithm that computes the depth and surface orientation simultaneously. In the proposed algorithm the local 3D surface is approximated by an arbitrary plane, not necessarily fronto-parallel, and a shape of a window is adaptively changed depending on the surface orientation. In order to avoid an exhaustive search for the surface orientation, we introduce a direct method to compute the surface orientation from the intensity gradients within the window. The method also copes with the affine intensity distortion between the corresponding regions. Proceeded by the theoretical formulation of the proposed algorithm, experimental results demonstrate that it effectively increases the accuracy of 3D shape recovery while being applicable to various images.

2 Background

In area-based stereo matching, there are at least two kinds of approaches to attenuate the influence caused by the phenomenon that the surface orientations locally deform the intensity patterns surrounding corresponding points.

One approach is to adaptively change the window size according to the amount of depth variation within the window. Okutomi[12] proposed the adaptive-window algorithm which selects an appropriate size of window by evaluating the local variation of both intensity and disparity. The algorithm can also deal with the depth discontinuities. However, since it requires the initial disparity estimates at numerous pixels within the window to select a suitable window size, those estimates need to be guessed rather accurately from the beginning. Unfortunately, the requirement is often too strict and unreasonable in terms of computational cost.

The other approach, which we also consider in this paper, is to compute the surface orientation as well as the depth. There have been several studies on computing the surface orientation from stereo vision. Gårding[16] used the windowed second moment matrix to estimate the linear spatial distortion. This technique allows a local estimate of the surface orientation to be computed directly from the local statistics of the left and right image intensity gradients. Jones and Malik[6] solved the same task by using a set of linear filters. Robert[10] investigated algorithms for evaluating the surface orientation without reconstructing an explicit metric representation of the scene. These methods, however, treat the problem of determining the surface orientation directly from images after establishing the correspondences and do not deal with the deformation problem of corresponding regions in the stereo matching. Devernay[7] proposed an enhanced correlation method that allows

a matching window to locally deform between a stereo image pair. Although this method shares a similar framework to ours, it represents the image deformation with respect to the derivatives of disparity, not to the surface orientation. Thereby, it is rather complex to utilize more than two images for a robust depth recovery because the disparities of different image pairs are not identical, nor are the derivatives of the disparities consequently. Mainome[14] tackled the deformation problem in phase-based stereo, but the scheme still needs the brute force search over the surface orientation in addition to the depth search.

Also in motion estimation or feature tracking, several researchers addressed the deformation problem of the intensity patterns. Rehg[8] used the polynomial deformation model to track a target object in image sequence. Fuh[3] presented the affine-deformation model for motion estimation. In these methods, however, apart from the expensive computational cost, it is extremely difficult to estimate the full six affine-deformation parameters including the relative camera motion as pointed out in [9, 11]. Several practical approaches have been therefore proposed. Shi[11] used the affine-deformation model to monitor the quality of the estimated motion. Bergen[9] utilized the deformation model to compute the planar surface flow assuming that the relative camera motion is already known. Manmatha[17] measured similarity transformation, i.e., a scale change and rotation by deforming the filter according to the local image deformation.

While these efforts are toward motion estimation or feature tracking, this paper addresses the deformation problem in stereo matching. We practically reduce the problem to that of finding three unknown parameters; one for depth and two for surface orientation. Directly obtaining the surface orientation from intensity gradients within the window, we solve for point correspondences taking into account the local surface orientation at a reasonable computational cost. In computing the surface orientation, we employ a more general model to deal with an affine intensity distortion between corresponding regions while most of conventional methods are based on the brightness consistency assumption. Further, since the proposed algorithm represents the local image deformation with respect to the surface orientation, it is straightforward to utilize multiple images as input and in fact we use three images to stabilize the depth estimation.

3 Local Image Deformation

As stated above the viewing geometry and the local surface orientations cause deformations of intensity patterns surrounding corresponding points. Strictly speaking, these deformations are represented by homography[10, 15]. However, since a depth variation within a window is generally small, they can be approximated by 2×2 affine-deformation matrices. This linear approximation enables a direct computation of the surface orientation. In this section, we analyze these deformations.

The stereo geometry is shown in Figure 1. We align the world coordinate with the first camera. Let π be a 3D plane, and $\boldsymbol{X} = (X, Y, Z)^\top$ and $\boldsymbol{X} + \boldsymbol{\Delta X} = (X + \Delta X, Y + \Delta Y, Z + \Delta Z)^\top$ be 3D points on π. The first camera projects 3D points $\boldsymbol{X}$ and $\boldsymbol{X} + \boldsymbol{\Delta X}$ onto 2D image points $\boldsymbol{x} = (x, y)^\top$ and $\boldsymbol{x} + \boldsymbol{\xi} = (x + \xi, y + \eta)^\top$ respectively. The perspective camera model defines these projections as

$$\boldsymbol{x} = \frac{F}{Z}\begin{bmatrix} X \\ Y \end{bmatrix},\ \boldsymbol{x} + \boldsymbol{\xi} = \frac{F}{Z + \Delta Z}\begin{bmatrix} X + \Delta X \\ Y + \Delta Y \end{bmatrix}, \tag{1}$$

where F is a focal length. Analogously, the second camera projects $\boldsymbol{X}$ and $\boldsymbol{X} + \boldsymbol{\Delta X}$

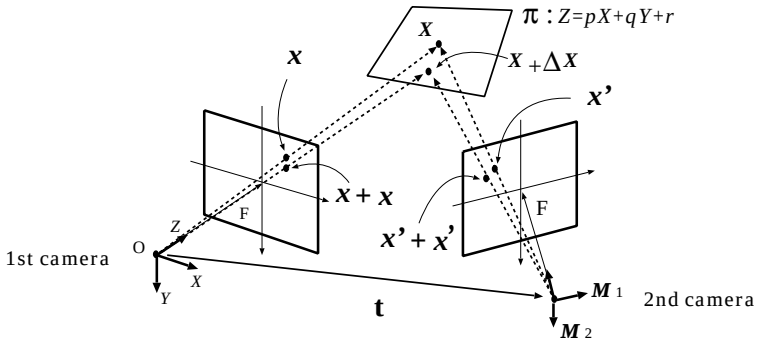


Figure 1: The stereo geometry

onto $\boldsymbol{x}' = (x', y')^\top$ and $\boldsymbol{x}' + \boldsymbol{\xi}' = (x' + \xi', y' + \eta')^\top$ respectively. Let $\boldsymbol{M}_1 = (m_{11}, m_{12}, m_{13})^\top$ and $\boldsymbol{M}_2 = (m_{21}, m_{22}, m_{23})^\top$ be respectively the unit vectors pointing along the scan-lines and the columns of the second camera image. The perspective camera model defines these projections as

$$\boldsymbol{x}' = \frac{F}{Z'} \begin{bmatrix} \boldsymbol{M}_1^\top (\boldsymbol{X} - \boldsymbol{t}) \\ \boldsymbol{M}_2^\top (\boldsymbol{X} - \boldsymbol{t}) \end{bmatrix}, \; \boldsymbol{x}' + \boldsymbol{\xi}' = \frac{F}{Z' + \Delta Z'} \begin{bmatrix} \boldsymbol{M}_1^\top (\boldsymbol{X} + \boldsymbol{\Delta X} - \boldsymbol{t}) \\ \boldsymbol{M}_2^\top (\boldsymbol{X} + \boldsymbol{\Delta X} - \boldsymbol{t}) \end{bmatrix}, \quad (2)$$

where $\boldsymbol{t} = (t_x, t_y, t_z)^\top$ is the translation vector from the world origin O to the origin of the second camera, and Z' and $Z' + \Delta Z'$ are respectively the depths of $\boldsymbol{X}$ and $\boldsymbol{X} + \boldsymbol{\Delta X}$ viewed from the second camera. Let the equation of the 3D plane π be $Z = pX + qY + r$. Since both $\boldsymbol{X}$ and $\boldsymbol{X} + \Delta \boldsymbol{X}$ exist on π, ΔZ is given by $\Delta Z = p\Delta X + q\Delta Y$. From this relationship and Equation(1) and (2), we derive

$$\boldsymbol{\xi}' = \boldsymbol{A}\boldsymbol{\xi} = (\boldsymbol{A}_c + \boldsymbol{A}_v)\boldsymbol{\xi}, \quad (3)$$

where $\boldsymbol{A} = \boldsymbol{A}_c + \boldsymbol{A}_v$,

$$\boldsymbol{A}_c = \frac{Z}{Z'} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \text{ and } \boldsymbol{A}_v = \frac{Z}{Z'} \begin{bmatrix} pm_{13} & qm_{13} \\ pm_{23} & qm_{23} \end{bmatrix}.$$

Here we have assumed $\Delta Z \ll Z$ and $\Delta Z' \ll Z$. Equation(3) represents the deformation of the corresponding area in the stereo images as shown in Figure 2. In the proposed algorithm we deform the window using this equation. The matrix $\boldsymbol{A}_c$ can be computed with the pose of the second camera $\boldsymbol{M} = (\boldsymbol{M}_1^\top, \boldsymbol{M}_2^\top)^\top$ and the scale factor $s = Z/Z'$. Since $\boldsymbol{M}$ is already known and s is also derived through the depth search, $\boldsymbol{A}_c$ is straightforwardly determined. On the other hand, the matrix $\boldsymbol{A}_v$ includes variable parameters depending on the image region, i.e., the local surface orientation (p, q). In order to deform

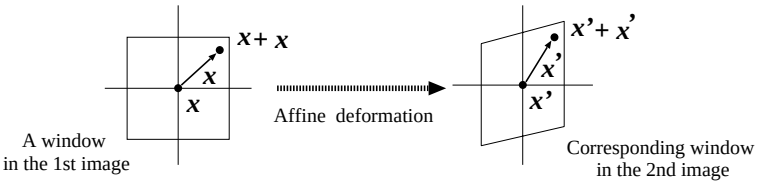


Figure 2: Window deformation

the window using Equation(3), we need to evaluate the surface orientation. Although $\boldsymbol{A}_v$ is conventionally ignored based on the preceding local fronto-parallel assumption, $p = q = 0$, the proposed algorithm takes it into account to improve the precision of matching.

4 Surface Orientation from Intensity Gradients

An exhaustive search for the surface orientation on the top of the depth search would make the computational cost extremely expensive. In this section, we introduce a direct method to compute the surface orientation from the local intensity gradients, that is inspired by the gradient-based optical flow computation[1]. In our method the following equation is employed[3],

$$af(\boldsymbol{x}_1 + \boldsymbol{\xi}) + b = g(\boldsymbol{x}_2 + \boldsymbol{\xi}'), \tag{4}$$

where $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are the image intensities of the image point $\boldsymbol{x} = (x, y)^\top$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are corresponding points, and $\boldsymbol{x}_1 + \boldsymbol{\xi}$ and $\boldsymbol{x}_2 + \boldsymbol{\xi}'$ are arbitrary corresponding points within the window centered around $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively, as shown in Figure 2. The parameter a is a scale factor introduced to account for the differences in image contrast, and b is a bias term modeling the possible differences in the mean intensity level. The parameters a and b are variable depending on the image region. Gradient-based approaches to compute optical flow usually use the brightness constancy assumption, so that $a = 1$ and $b = 0$ in Equation(4). Thus, Equation(4) formulates a more general model which deals with the affine intensity distortion between the corresponding regions. This extension is important for the stereo matching.

Using Equation(3) and Taylor expansion, the right-hand side of Equation(4) becomes

$$g = g(\boldsymbol{x}_2 + \boldsymbol{A}_c\boldsymbol{\xi} + \boldsymbol{A}_v\boldsymbol{\xi}) = g(\boldsymbol{x}_2 + \boldsymbol{A}_c\boldsymbol{\xi}) + \nabla g(\boldsymbol{x}_2 + \boldsymbol{A}_c\boldsymbol{\xi})\boldsymbol{A}_v\boldsymbol{\xi}, \tag{5}$$

where $\nabla g = (g_x, g_y)$ is the intensity gradient. Here the higher order terms of $\boldsymbol{A}_v\boldsymbol{\xi}$ are ignored. Using this expansion, Equation(4) can be rewritten as

$$\left[\ \Gamma(\boldsymbol{\xi})\xi \ \ \Gamma(\boldsymbol{\xi})\eta \ \ -f(\boldsymbol{x}_1 + \boldsymbol{\xi}) \ \ -1\ \right]\boldsymbol{\alpha} = -g(\boldsymbol{x}_2 + \boldsymbol{A}_c\boldsymbol{\xi}), \tag{6}$$

where $\boldsymbol{\alpha} = (p, q, a, b)^\top$ and $\Gamma(\boldsymbol{\xi}) = s\{m_{13}g_x(\boldsymbol{x}_2 + \boldsymbol{A}_c\boldsymbol{\xi}) + m_{23}g_y(\boldsymbol{x}_2 + \boldsymbol{A}_c\boldsymbol{\xi})\}$. Since Equation(6) holds for any point within the window, a sufficient number of equations to estimate the unknown parameters of $\boldsymbol{\alpha}$ are acquired by picking up arbitrary points. Hence, the surface orientation can be computed using a minimum of two images while the use of multiple images makes the estimation more robust.

Let the third image be $h(\boldsymbol{x})$ and the corresponding point be $\boldsymbol{x}_3$. The following equation is employed, $a'f(\boldsymbol{x}_1 + \boldsymbol{\xi}) + b' = h(\boldsymbol{x}_3 + \boldsymbol{\xi}'')$. Analogously, we obtain

$$\left[\ \Gamma'(\boldsymbol{\xi})\xi \ \ \Gamma'(\boldsymbol{\xi})\eta \ \ -f(\boldsymbol{x}_1 + \boldsymbol{\xi}) \ \ -1\ \right]\boldsymbol{\beta} = -h(\boldsymbol{x}_3 + \boldsymbol{A}'_c\boldsymbol{\xi}), \tag{7}$$

where $\boldsymbol{\beta}$=$(p, q, a', b')^\top$ and $\Gamma'(\boldsymbol{\xi})$=$s'\{m'_{13}h_x(\boldsymbol{x}_3 + \boldsymbol{A}'_c\boldsymbol{\xi}) + m'_{23}h_y(\boldsymbol{x}_3 + \boldsymbol{A}'_c\boldsymbol{\xi})\}$ [1]. Since at any point $\boldsymbol{\xi} = \boldsymbol{\xi}_i (i = 1, 2, ..., N)$ within the window satisfies both Equation(6) and (7), the unknown parameters $\boldsymbol{\gamma} = (p, q, a, b, a', b')^\top$ can be estimated. As this method uses only the intensity gradients within the window, it enables the depth estimation with

[1] Note that the primes indicate these parameters are for the third image $h(x)$.

a locally affine-deformable window depending on the surface orientation at a reasonable computational cost.

It is also possible to refine the surface orientation estimate with iteration. Let (p_k, q_k) be a current estimate of (p, q) during the *k*-th iteration and $(\Delta p, \Delta q)$ be an incremental estimate. The preceding window deformation matrices become

$$\boldsymbol{A}_c = s \begin{bmatrix} m_{11} + p_k m_{13} & m_{12} + q_k m_{13} \\ m_{21} + p_k m_{23} & m_{22} + q_k m_{23} \end{bmatrix}, \quad \boldsymbol{A}_v = s \begin{bmatrix} \Delta p m_{13} & \Delta q m_{13} \\ \Delta p m_{23} & \Delta q m_{23} \end{bmatrix}.$$

Since $(\Delta p, \Delta q)$ can be computed analogously, we can update the surface orientation estimate by $(p_{k+1}, q_{k+1}) = (p_k, q_k) + (\Delta p, \Delta q)$.

5 Stereo Matching Algorithm

In this section, we describe our stereo matching algorithm based on the analysis in the previous sections. Although it is possible to recover the 3D shape from at least two images, we utilize three images to reduce the ambiguity of correspondence and to estimate the surface orientation robustly as described in Section 4. Given three images captured from different viewpoints, we must solve the viewing geometry at the beginning. For simplification we assume the orthographic camera model. Under orthography the viewing geometry can be solved using SVD[5] given the correct corner correspondences. Since some incorrect correspondences may be inevitably involved, we employ a random sampling consensus(RANSAC) technique to remove the mismatches[4].

Our concern now is in stereo matching, namely how to recover the 3D shape using the geometric constraint. In the following we summarize the stereo matching algorithm with a locally affine-deformable window.

1. Let $\boldsymbol{x}$ be the image point where we want to compute the depth[2]. Compute the initial depth estimate Z_0 of $\boldsymbol{x}$, assuming the 3D surface is locally frontal.
2. Assuming the depth is Z_0, compute the surface orientation (p, q) of $\boldsymbol{x}$.
3. Compute the local affine-deformation matrices $\boldsymbol{A}(j)(j = 2, 3)$ by Equation(3) and evaluate the consistency $C(Z)$ of the depth Z. We define the consistency $C(Z)$ as the sum of two normalized cross correlation measures taking into account the window deformations.
4. At each depth Z within the search range $(Z_0 - \Delta Z < Z < Z_0 + \Delta Z)$, compute the consistency $C(Z)$. The depth Z^* which gives the maximum value of $C(Z)$ is basically regarded as the correct depth of $\boldsymbol{x}$.

6 Experiments

We present some experiments in which we have applied our stereo matching algorithm to real images.

First, we compare the performance of the proposed algorithm with that of the conventional one which assumes the 3D surface to be locally fronto-parallel. For precise evaluation of the measurement error, we use a simple object whose shape is already known as shown in Figure 3. This is a rubber baseball whose radius R is about 3.5cm. We use a CCD camera with a 25mm lens whose image resolution is 480×480 pixels. The distance from the camera to the object is about 50cm.

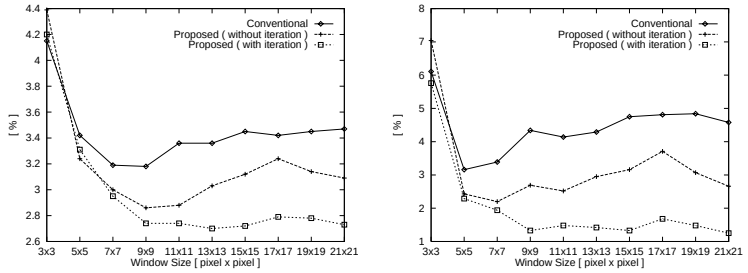
[2] Strictly speaking, the "depth" cannot be recovered under orthography. But for convenience we use the term of "depth" in this paper.

Figure 4(a) shows the average errors in the recovered depths with various window sizes by the conventional algorithm and the proposed one (with and without iteration in computing the surface orientation). The error is normalized by the radius R. With a 3×3 window, the shape estimation is poor because such a small window does not include sufficient intensity variation to identify the corresponding point. In order to eliminate the ambiguity of correspondence, it is necessary to use larger windows. With 5×5 or larger windows, our algorithm always gives better shape estimation than does the conventional one. Moreover, the estimates with iteration in computing the surface orientation are better than those without iteration. Figure 4(b) shows the percentage of the reference points which have gross depth errors. We define the gross depth error as above 20 pixels(about 4mm). The tendency is similar to the case with the average error and it is clear that our algorithm realizes a smaller number of gross errors. Figure 5 shows the recovered shape by the proposed algorithm with 11×11 window.

We have also applied the proposed algorithm to various images of, such as human face and human hand. Figure $6\sim10$ show the input images and the recovered 3D shapes. The distance from the camera to the object is about 50cm. Figure 8 shows the 3D shapes of the nose in "Face" obtained by the conventional algorithm and the proposed one.

It is not an easy task to recover these shapes because the input images are only slightly

Figure 3: "Ball" images (480 $\times$ 480)



(a) Average relative error (b) Percentage of gross error

Figure 4: Comparison of the precision

Figure 5: Recovered shape of "Ball" by the proposed algorithm

textured. Nevertheless, the precise shapes are acquired and these results show that our algorithm works for a wide variety of images. Figure 8 exemplifies that the proposed algorithm enables more accurate recovery of the 3D shape than does conventional one. The proposed algorithm successfully recovers the flat structure of the circle area in Figure 8(a), while the conventional one fails. The parts pointed by the arrows in Figure 8 (b) and (c) correspond to the circle area in (a).

Figure 6: “Face” images

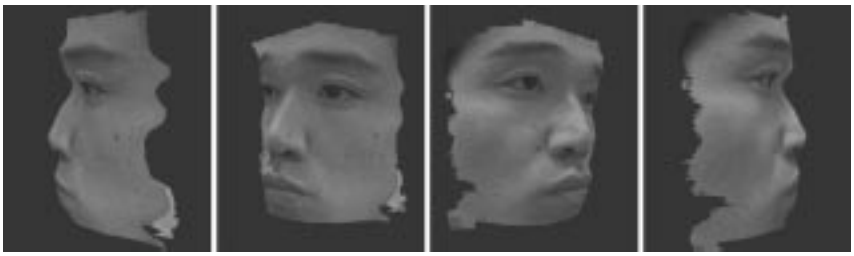
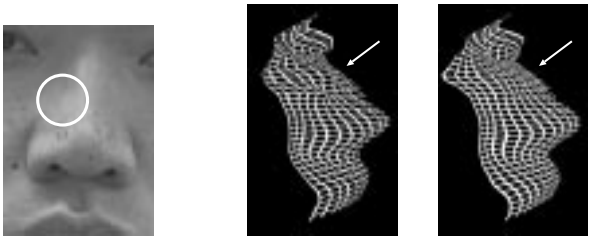


Figure 7: Recovered shape of “Face”



(a) (b) (c)

Figure 8: (a)Nose area in “Face”. (b) Recovered shape by the conventional algorithm. (c) Recovered shape by the proposed algorithm

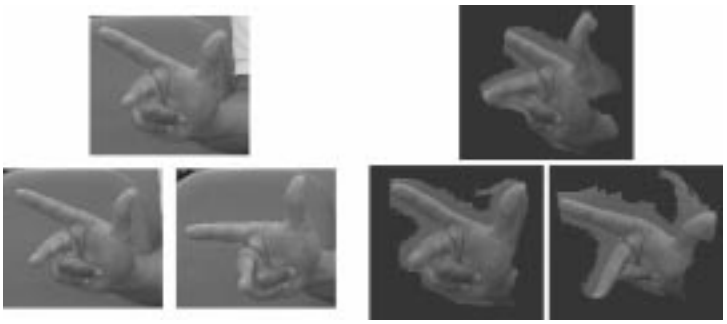


Figure 9: “Hand1” images (left) and recovered shape (right)

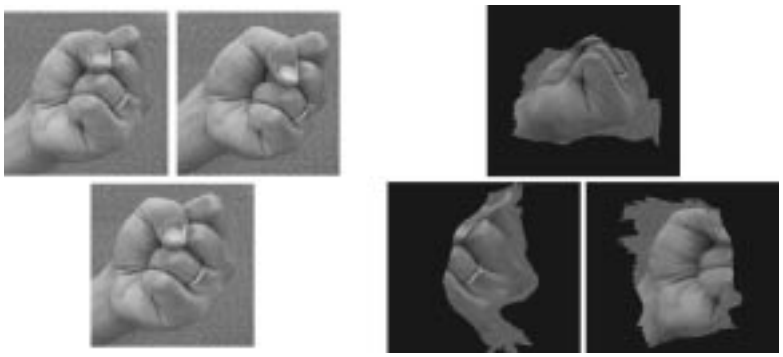
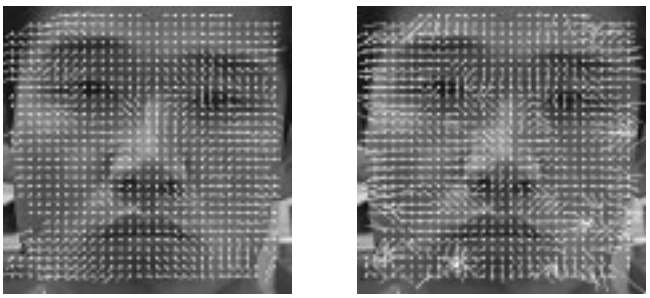


Figure 10: “Hand2” images (left) and recovered shape (right)

The proposed algorithm produces not only the depth but also the surface orientation. Figure 11(a) shows the generated surface orientations of “Face” by our algorithm whereas Figure 11(b) shows those by locally fitting a plane on the depth map. The surface orientations derived from the depth map are noisy because of direct influence by the errors included in the depth estimations. On the other hand, those recovered directly from intensity gradients are more reasonable. The result indicates that the proposed algorithm is



(a) Proposed (b) From depth map

Figure 11: Recovered surface orientations of “Face”

more appropriate also in terms of recovering the surface orientation.

7 Summary and Conclusions

In this paper we have proposed a new stereo matching algorithm which computes the depth and surface orientation simultaneously. In our algorithm, the 3D surface is locally approximated by a plane whose surface orientation is arbitrary while the conventional ones implicitly assume the 3D surface to be locally fronto-parallel. The proposed algorithm locally deforms a window according to the surface orientation which is directly recovered from intensity gradients within the window. The experimental results have demonstrated a clear advantage of our algorithm over conventional ones and it is applicable to a wide variety of images. In future work, we will use hierarchical approaches to improve the efficiency and to treat the depth discontinuities. We are also planning to combine the proposed algorithm with the techniques of Structure-From-Motion under more general projection models[2, 13] in order to alleviate the limitations to the scene and relative camera motion.

References

[1] B.K.P.Horn and B.G.Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[2] B.Triggs. Factorization methods for projective structure and motion. In *Proc.CVPR*, pages 845–851, 1996.

[3] C.Fuh and P.Maragos. Motion displacement estimation using an affine model for matching. *Optical Engineering*, 30(7):881–887, 1991.

[4] C.S.Wiles, A.Maki, N.Matsuda, and M.Watanabe. Hyper-patches for 3d model acquisition and tracking. In *Proc.CVPR*, pages 1074–1080, 1997.

[5] C.Tomasi and T.Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9:2:137–154, 1992.

[6] D.G.Jones and J.Malik. Determining three-dimensional shape from orientation and spatial frequency disparities. In *Proc. 2nd ECCV*, pages 661–669, 1992.

[7] F.Devernay and O.Faugeras. Computing differential properties of 3D shapes from stereoscopic images without 3D models. Technical report 2304, INRIA, 1994.

[8] J.M.Rehg and A.P.Witkin. Visual tracking with deformation models. In *Proc.International Conference on Robotics and Automation*, pages 844–850, 1991.

[9] J.R.Bergen, P.Anandan, K.J.Hanna, and R.Hingorani. Heirachical model-based motion estimation. In *Proc. 2nd ECCV*, pages 237–252, 1992.

[10] J.Robert and M.Hebert. Deriving orientation cues from stereo images. In *Proc. 3rd ECCV*, pages 377–388, 1994.

[11] J.Shi and C.Tomasi. Good features to track. In *Proc.CVPR*, pages 593–600, 1994.

[12] M.Okutomi and T.Kanade. A locally adaptive window for signal matching. *IJCV*, 7:2:143–162, 1992.

[13] M.Pollefeys, R.Koch, and L.V.Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th ICCV*, pages 90–95, 1998.

[14] M.W.Maimone and S.A.Shafer. Modeling foreshortening in stereo vision using local spatial frequency. In *Proc.IROS*, pages 519–524, 1995.

[15] P.Torr, A.W.Fitzgibbon, and A.Zisserman. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In *Proc. 6th ICCV*, pages 485–491, 1998.

[16] J.Gårding and T.Lindeberg. Direct estimation of local surface shape in a fixating binocular vision system. In *Proc. 3rd ECCV*, pages 365–376, 1994.

[17] R.Manmatha. Measuring the affine transform using gaussian filters. In *Proc. 3rd ECCV*, pages 159–164, 1994.

[18] U.R.Dhond and J.K.Aggarwal. Structure from stereo - a review. *Trans. Systems, and Man Cybernetics*, 19:1489–1510, 1989.