A SUB-PIXEL STEREO CORRESPONDENCE TECHNIQUE
BASED ON 1D PHASE-ONLY CORRELATION

Takuma Shibahara[†], Takafumi Aoki[†]

†Tohoku University
Graduate School of Information Sciences
Sendai-shi 980-8579, Japan
E-mail: aoki@ecei.tohoku.ac.jp

Hiroshi Nakajima[††], Koji Kobayashi[††]

††Yamatake Corporation
Isehara-shi 259-1195, Japan

ABSTRACT

This paper presents a technique for high-accuracy correspondence search between two rectified images using 1D Phase-Only Correlation (POC). The correspondence search between stereo images can be reduced to 1D search through image rectification. However, we usually employ block matching with 2D rectangular image blocks for finding the best matching point in the 1D search. We propose the use of 1D POC (instead of 2D block matching) for stereo correspondence search. The use of 1D POC makes possible significant reduction in computational cost without sacrificing reconstruction accuracy compared with the 2D POC-based approach. Also, the resulting reconstruction accuracy is much higher than those of conventional stereo matching techniques using SAD (Sum of Absolute Differences) and SSD (Sum of Squared Differences) combined with sub-pixel disparity estimation.

Index Terms— 3D measurement, stereo vision, stereo correspondence, sub-pixel image matching, phase-based image matching, phase-only correlation

1. INTRODUCTION

Recently the demand of high-accuracy 3D measurement is rapidly growing in a variety of computer vision applications [1]. Existing 3D measurement techniques are classified into two major types — active and passive. In general, active measurement employs structure illumination (structure projection, phase shift, moire topography, etc.) or laser scanning, which is not desirable in many applications. On the other hand, passive 3D measurement techniques based on stereo vision have the advantages of simplicity and applicability, since such techniques require simple instrumentation. However, poor reconstruction quality still remains as a major issue for passive 3D measurement, due to the difficulty in finding accurate correspondence between stereo images [2].

The most common stereo correspondence techniques employ Sum of Absolute Differences (SAD) or Sum of Squared Differences (SSD), where corresponding points between stereo images can be obtained by minimizing SAD or SSD in area-based block matching [3, 4]. Although SAD and SSD exhibit low computational cost, a major drawback is their low accuracy. Recently, sub-pixel block matching techniques using SAD and SSD have been investigated [4], but the obtained accuracy is not sufficient in some applications.

On the other hand, image matching methods using 2D Phase-Only Correlation (POC) [1] exhibit much better matching performance

---

[1]"Phase-only correlation" is sometimes called the "phase correlation."

than the methods using SAD and SSD in general [5, 6, 7]. The authors have already developed POC-based passive 3D measurement system, whose accuracy is comparable with those of projector-based active 3D measurement systems [8, 9]. A drawback of POC-based approach is its high computational cost in evaluating the 2D POC for correspondence search, which limits the area of applications.
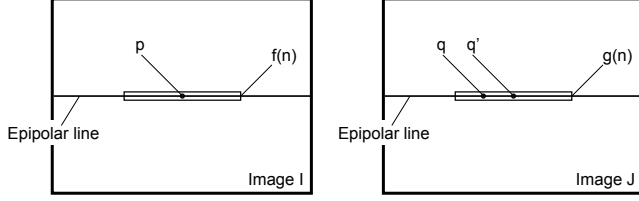
Addressing this problem, in this paper, we propose a technique for high-accuracy correspondence search between two rectified images using 1D version of POC. The correspondence search between stereo images can be reduced to 1D search through image rectification. However, conventional approach is to employ block matching with 2D rectangular image blocks for finding the best matching point within 1D search interval. In this paper, on the other hand, we propose the use of 1D POC (instead of 2D block matching) for stereo correspondence search. The use of 1D POC makes possible significant reduction in computational cost without sacrificing reconstruction accuracy, compared with the 2D POC-based approach. Also, the resulting reconstruction accuracy is much higher than those of conventional stereo matching techniques using SAD (Sum of Absolute Differences) and SSD (Sum of Squared Differences) combined with sub-pixel disparity estimation [4]. A set of experiments demonstrate that the stereo vision system employing the proposed technique can measure 3D surfaces of free-form objects with sub-mm accuracy.

2. 1D PHASE-ONLY CORRELATION

This section defines the 1D POC function and a set of techniques for high-accuracy image matching. Let I and J be rectified stereo images as illustrated in Fig. 1. Given a reference point p in the image I, the problem is to find the corresponding point q in the image J. In the image I, we first extract the 1D image signal $f(n)$ centered at the reference point p along the epipolar line. Similarly, in the image J, we extract the 1D image signal $g(n)$ centered at q' — the initial estimate for the true corresponding point q. The points q and q' should be on the common epipolar line corresponding to p. Let n ($\in \{-M, -(M-1), \cdots, 0, \cdots, (M-1), M\}$) be the discrete spatial index for the 1D image signals $f(n)$ and $g(n)$, where M is a positive integer. The signal length N is given by $N = 2M + 1$. Note that we assume here the sign symmetric index range $\{-M, \cdots, M\}$ for mathematical simplicity. The discussion could be easily generalized to non-negative index ranges with power-of-two signal length.

The 1D Discrete Fourier Transforms (1D DFTs) of $f(n)$ and

Fig. 1. Rectified stereo images.

Fig. 2. Set of 1D image signals used for averaging 1D POC functions.

$g(n)$ are given by

$$F(k) = \sum_{n=-M}^{M} f(n)W_N^{kn} = A_F(k)e^{j\theta_F(k)}, \quad (1)$$

$$G(k) = \sum_{n=-M}^{M} g(n)W_N^{kn} = A_G(k)e^{j\theta_G(k)}, \quad (2)$$

where $k = -M \cdots M$ and $W_N = e^{-j\frac{2\pi}{N}}$. A_F and A_G are amplitude components, and $e^{j\theta_F(k)}$ and $e^{j\theta_G(k)}$ are phase components. The cross-phase spectrum $R(k)$ is defined as

$$R(k) = \frac{F(k)\overline{G(k)}}{|F(k)\overline{G(k)}|} = e^{j\theta(k)}, \quad (3)$$

where $\overline{G(k)}$ denotes the complex conjugate of $G(k)$ and $\theta(k) = \theta_F(k) - \theta_G(k)$. The 1D POC function $r(n)$ between $f(n)$ and $g(n)$ is the 1D Inverse DFT (1D IDFT) of $R(k)$ and is given by

$$r(n) = \frac{1}{N} \sum_{k=-M}^{M} R(k)W_N^{-kn}. \quad (4)$$

In the following, we derive the analytical peak model for the 1D POC function between the same signals that are minutely displaced with each other. Now consider $f_c(x)$ as a 1D image signal defined in continuous space with real-number index x. Let δ represents minute (sub-pixel) displacement of $f_c(x)$. So, the displaced 1D image signal can be represented as $f_c(x - \delta)$. Assume that $f(n)$ and $g(n)$ are spatially sampled signals of $f_c(x)$ and $f_c(x - \delta)$, respectively, and are defined as
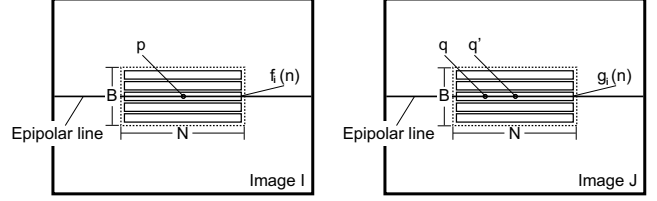
$$f(n) = f_c(x)|_{x=nT}, \quad (5)$$
$$g(n) = f_c(x - \delta)|_{x=nT}, \quad (6)$$

where T is the spatial sampling interval, and index range is given by $n = -M, \cdots, M$. For simplicity, we assume $T = 1$. The POC function $r(n)$ between $f(n)$ and $g(n)$ is given by

$$r(n) \simeq \frac{\alpha}{N} \frac{\sin\{\pi(n+\delta)\}}{\sin\{\frac{\pi}{N}(n+\delta)\}}, \quad (7)$$

where $\alpha = 1$. The above Eq. (7) represents the shape of the peak for the 1D POC function between the same 1D image signals that are minutely displaced with each other. This equation gives a distinct sharp peak. (When $\delta = 0$, the 1D POC function $r(n)$ becomes the Kronecker delta function.) We can show that the peak value α decreases (without changing the function shape itself), when small noise components are added to the original images. Hence, we assume $\alpha \leq 1$ in practice. The peak position $n = -\delta$ of the 1D POC function reflects the displacement between the two 1D image signals.

Thus, we can compute the displacement δ between extracted signals $f(n)$ and $g(n)$ by estimating the true peak position of the 1D POC function $r(n)$. Then, the corresponding point q for the reference point p is determined from q' and δ as

$$q = q' + (\delta, 0), \quad (8)$$

where q and q' in this equation are regarded as the coordinate vectors of the true corresponding point q and its initial estimate q', respectively.

Listed below are important techniques for improving the accuracy of 1D image matching for sub-pixel correspondence search.

(i) Function fitting for high-accuracy estimation of peak position
We use Eq. (7) — the closed-form peak model of the POC function — directly for estimating the peak position by function fitting. By calculating the POC function, we can obtain a data of $r(n)$ for each discrete index n. It is possible to find the location of the peak that may exist between image pixels by fitting the function Eq. (7) to the calculated data array around the correlation peak, where α and δ are fitting parameters.
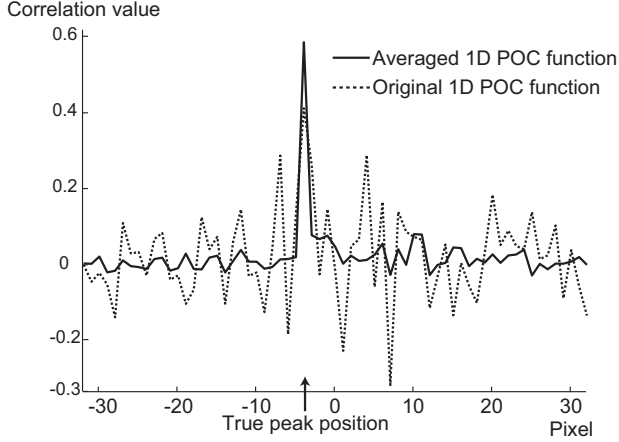
(ii) Windowing to reduce boundary effects
Due to the DFT's periodicity, a signal can be considered to "wrap around" at an edge, and therefore discontinuities, which are not supposed to exist in real world, occur at every border in 1D DFT computation. We reduce the effect of discontinuity at signal border by applying 1D window function to 1D image signals. For this purpose, we employ 1D Hanning window.

(iii) Spectral weighting for reducing aliasing and noise effects
For natural images, typically the high frequency components may have less reliability (low S/N) compared with the low frequency components. We could improve the estimation accuracy by applying a low-pass-type weighting function to 1D POC function in frequency domain and eliminating the high frequency components with low reliability. For this purpose, we use the Gaussian-type spectral weighting function. The peak model Eq. (7) for function fitting should be modified correspondingly.

(iv) Averaging 1D POC functions to improve peak-to-noise ratio
When image quality is poor, a single 1D POC function is not sufficient for estimating accurate correspondence q due to degraded Peak-to-Noise Ratio (PNR). We can improve PNR by averaging a set of 1D POC functions evaluated at distinct positions around p and q'. Figure 2 illustrates a typical situation. We extract B distinct 1D image signals $f_i(n)$ $(i = 1, 2, \cdots, B)$ around the reference point p in the image I. Similarly, we extract 1D image signals $g_i(n)$ $(i = 1, 2, \cdots, B)$ around the initial estimate q' in the image J. Then, we compute the B distinct 1D POC functions $r_i(n)$ between $f_i(n)$ and $g_i(n)$. By taking the average of $r_i(n)$ for $i = 1, 2, \cdots, B$, we have the overall correlation surface $r(n)$ with significantly improved PNR. Figure 2 illustrates a typical case of $B = 5$, which can be easily generalized to arbitrary arrangement of 1D image signals. Figure 3 shows an example of PNR improvement through averaging.

Fig. 3. Averaging 1D POC functions to improve Peak-to-Noise Ratio (PNR).

Table 1. Computational cost for finding a single corresponding point.

	ADD	MUL	DIV	SQRT
1D POC (11x32 pixels)	12,167	11,010	709	352
2D POC (32x32 pixels)	95,667	72,340	2,058	1024
SAD (32x32 pixels)	32,754	1	1	0
SSD (32x32 pixels)	39,925	20,487	2	0

Camera: Adimec-1000m/D
10 bits digital resolution
monochrome
1004 x 1004 pixels

Lens: µTRON, FV1520
15 mm focal length

Image grabber: Coreco Imaging
X64-CL-DUAL-32M

Stereo baseline: 46 mm

Measurement range: 400 ~ 600 mm

(a)   (b)

Fig. 4. Stereo vision system: (a) stereo camera head, and (b) system specification.

(v) Coarse-to-fine strategy for robust correspondence search
In our stereo matching algorithm based on 2D POC, we have adopted a coarse-to-fine strategy using image pyramids for robust correspondence search [9]. The reason is that dense stereo correspondence requires matching of smaller image blocks, while the accuracy and robustness of POC-based image matching degrade significantly as the image size decreases. The coarse-to-fine approach is highly effective for solving this problem when combined with 2D POC. Our observation shows that the same problem occurs also for 1D POC when N becomes small, e.g., $N = 32$. Hence, we have developed a 1D version of the coarse-to-fine matching algorithm, where we employ multi-resolution image pyramid (with 3 layers in typical applications) for robust correspondence search.

3. EXPERIMENTS AND DISCUSSION

The proposed sub-pixel stereo matching technique allows us to implement a high-accuracy passive 3D measurement system, whose accuracy may be comparable with those of projector-based active 3D measurement systems. In this section, we evaluate the performance of the proposed technique in terms of computational cost and accuracy. We compare four different techniques using (i) 1D POC, (ii) 2D POC [9], (iii) SAD [4] and (iv) SSD [4], where each technique is equipped with sub-pixel stereo matching capability.

3.1. Computational Cost

We evaluate the amount of computation required for finding a single corresponding point q within a fixed size of 1D search interval. The system parameters for the methods (i)–(iv) are optimized through actual 3D measurement experiments as described in the section 3.2. For (i) 1D POC, we assume that the length of 1D image signal is $N = 32$ and number of 1D image signals to be averaged is $B = 11$. For (ii) 2D POC, (iii) SAD and (iv) SSD, we assume that the matching block size is $N_1 \times N_2 = 32 \times 32$. The length of 1D search interval (i.e., maximum disparity) is 16 pixels. (Note that we assume the use of coarse-to-fine strategy based on image pyramids.)

Table 1 summarizes the computational cost for finding a single corresponding point, where "ADD", "MUL", "DIV" and "SQRT" denote the number of additions, multiplications, divisions and square

roots, respectively. By using 1D POC, significant reduction in computational cost is expected in comparison with 2D POC, where the number of additions/multiplications can be reduced to 1/7. Also, the amount of basic arithmetic operations is comparable with the methods using SAD and SSD.

3.2. Accuracy of 3D Measurement

We carried out a set of experiments for evaluating the accuracy and quality of 3D measurement. Figure 4 shows the stereo vision system used in our experiments, where two parallel cameras form a narrow-baseline stereo pair with baseline 46 mm. The system parameters for image matching are $N = 32$ and $B = 11$ for 1D POC, and $N_1 \times N_2 = 32 \times 32$ for 2D POC, SAD and SSD. All the techniques described in the Sect. 2 are employed, where the number of layers for coarse-to-fine search is 3.

At first, we evaluate the accuracy of 3D reconstruction using the reference object of geometrically regular shape — a solid sphere of radius 108.45 mm. The distance between the camera head and the reference objects is around 500 mm. In order to evaluate measurement accuracy for the solid sphere, we generate a best fitted sphere for the measured points by the least-squares algorithm. Table 2 compares the errors in 3D measurement by (i) 1D POC, (ii) 2D POC, (iii) SAD and (iv) SSD. This result shows that the proposed sub-pixel correspondence technique contributes to reducing the RMS (Root Mean Square) error and the maximum error, significantly.

In addition, a human face — a typical example of free form objects — is measured to demonstrate the capability of high-quality dense 3D reconstruction. Figure 5 (a) shows stereo images captured from the stereo cameras. Figures 5 (b)–(e) compare the quality of 3D surfaces produced by (i) 1D POC, (ii) 2D POC, (iii) SAD and (iv) SSD. The POC-based techniques (i) and (ii) can successfully reconstruct the smooth surface of the human face. The methods using SAD and SSD tend to produce matching errors and their quality is not high. To the best of the authors' knowledge, the quality of 3D re-

Table 2. Errors [mm] in 3D measurement of a sphere object.

|	1D POC	2D POC	SAD	SSD
RMS error	0.52	0.51	1.65	1.09
Max. error	1.92	2.09	25.68	24.06

construction using POC-based methods seems to be one of the best that is available with passive 3D measurement techniques reported to date. The result of this paper clearly suggests a potential possibility of our proposed approach to be widely used in many computer vision applications.

4. CONCLUSION

This paper presents a technique for high-accuracy correspondence search between two rectified images using 1D POC. The use of 1D POC in 1D correspondence search makes possible to significantly reduce computational cost without sacrificing reconstruction accuracy compared with the conventional 2D POC-based approach. Through some experimental evaluations, we demonstrate that the stereo vision system employing the proposed technique achieves sub-mm ($\sim$ 0.48 mm) accuracy in 3D measurement.

5. REFERENCES

[1] M. Petrov, A. Talapov, T. Robertson, A. Lebedev, A. Zhilyaev, and L. Polonskiy, "Optical 3D digitizers: Bringing life to the virtual world," *IEEE CG&A*, vol. 18, no. 3, pp. 28–37, May/Jun 1998.

[2] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 1, pp. 7–42, Apr. 2002.

[4] M. Shimizu and M. Okutomi, "Sub-pixel estimation error cancellation on area-based matching," *IJCV*, vol. 63, no. 3, pp. 207–224, July 2005.

[5] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," *Proc. Int. Conf. on Cybernetics and Society*, pp. 163–165, 1975.

[6] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. Fundamentals*, vol. E86-A, no. 8, pp. 1925–1934, Aug. 2003.

[7] K. Takita, M. A. Muquit, T. Aoki, and T. Higuchi, "A sub-pixel correspondence search technique for computer vision applications," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1913–1923, Aug. 2004.

[8] N. Uchida, T Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "3D face recognition using passive stereo vision," *Proc. of the 2005 IEEE ICIP*, pp. II–950–II–953, Sept. 2005.

[9] M. A. Muquit, T Shibahara, and T. Aoki, "A high-accuracy passive 3D measuremet system using phase-based image matching," *IEICE Trans. Fundamentals*, vol. E89-A, no. 3, pp. 686–697, Mar. 2006.

(a)

(b)

(c)

(d)

(e)
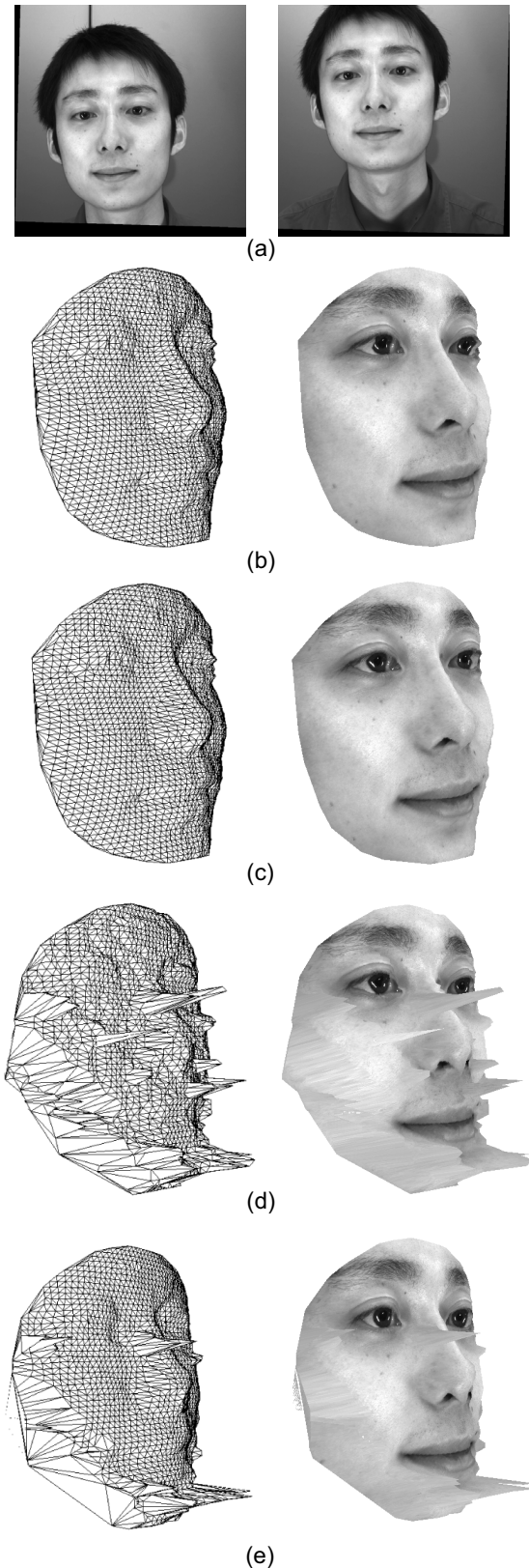
Fig. 5. Reconstructed 3D face data: (a) stereo images, and reconstructed 3D data by (b) 1D POC, (c) 2D POC, (d) SAD and (e) SSD.