

On Phase Based Stereo Matching and Its Related Issues

András Rövid, Takeshi Hashimoto

Abstract—The paper focuses on the problem of the point correspondence matching in stereo images. The proposed matching algorithm is based on the combination of simpler methods such as normalized sum of squared differences (NSSD) and a more complex phase correlation based approach, by considering the noise and other factors, as well. The speed of NSSD and the preciseness of the phase correlation together yield an efficient approach to find the best candidate point with sub-pixel accuracy in stereo image pairs. The task of the NSSD in this case is to approach the candidate pixel roughly. Afterwards the location of the candidate is refined by an enhanced phase correlation based method which in contrast to the NSSD has to run only once for each selected pixel.

Keywords—Stereo matching, Sub-pixel accuracy, phase correlation, SVD, NSSD.

I. INTRODUCTION

NOWADAYS the utilization of visual information plays crucial role in many applications, e.g. navigation, safety, security, medical systems, etc., but still with many open problems. It is well known that based on the information contained in digital images different type of objects and patterns can be detected, recognized and tracked, furthermore, in case of utilizing stereo images 3D measurement can be performed and thus the shape of objects or that of the whole scene can be reconstructed and modeled. There are many applications where the image processing plays crucial role; however, in this paper we are going to focus on the problematic related to stereo vision based 3D measurement, especially on the correspondence matching problem.

The field of stereo vision represents an active research area, many methods and algorithms have been developed in the last decades aiming to solve the stereo matching problem. It still represents an open issue thus new approaches are highly welcome. In the followings let us briefly mention some methods related to matching problem. The matching problem can easily be resolved by projecting a given type of coding pattern onto the surface of the target, however such approaches (first of all the projector-camera based systems due to relatively low intensity of the illumination or systems based on infrared light projection) are usually efficient for indoor applications only, i.e. where the lighting conditions are suitable for being able to unambiguously identify the pixels through pattern detection and decoding. Although 3D measurement systems based on laser light illumination are suitable for both indoor and outdoor applications, they have still limitations in

acquiring moving objects. The approaches mentioned until now - due to utilization of some type of external pattern - belong to the category of active measurement devices. With the help of these devices high accuracy measurements can be performed (few dozen of microns), thus they are highly welcome in reverse engineering related applications [1]-[4]. The second category relates to systems which are utilizing only images taken from different camera positions without any external pattern projection. In this case the point correspondence matching, i.e. finding the corresponding point to a given one chosen from the reference image represents the most crucial task. Many methods have been proposed for correspondence matching, which can be split into two main categories, i.e. into local and global searching methods. Local methods are based on searching the candidate point through calculating some kind of similarity metric (normalized cross correlation (NCC), sum of squared differences (SSD), sum of absolute differences (SAD), etc.) between two patterns. The pattern is usually considered over a rectangular window sliding along a scan-line (epipolar line). These methods estimate each corresponding pairs (the source image point and its corresponding one) independently, therefore they are called local. Many methods have been proposed for this problem, as for example methods based on selective multiple windows or changing the window size [5],[6].

On the other hand the global methods are attempting to minimize a global cost function (considering the piece-wise smooth property of disparities) usually composed from a data term and a smoothness term (discontinuity preserving). Such energy functions are often difficult to minimize (for example minimizing the Potts energy is NP-hard)[7], therefore only locally optimal solutions are utilized. As examples of this category graph-cut based methods can be emphasized, which can yield suitable results by finding the minimum cut of a special-purpose graph[7],[8]. The mean field annealing together with the dynamic programming and belief propagation, etc. are another popular approaches to find a particular solution to the above mentioned minimization problem [9],[11],[10].

In the present paper a local matching approach for point correspondence matching is introduced based. Although the computational complexity of the method (depending on the window size) is relatively high, many of its components can be run simultaneously.

The paper is organized as follows: In Section II the basic principle of the proposed method is described while in Sections III and IV a detailed analysis of the proposed approach can be followed. In Section V model evaluation and experimental results are discussed. Finally Section VI reports conclusions.

András Rövid is with the John von Neumann Faculty of Informatics, Óbuda University, Budapest, Bécsi út 96/B, Hungary, (e-mail: rovid.andras@nik.uni-obuda.hu).

Takeshi Hashimoto is with the Department of Electrical and Electronics Engineering, Shizuoka University, 5-1, 3-chome Johoku, Naka-ku, Hamamatsu, 432-8561, Japan, (e-mail: tethash@ipc.shizuoka.ac.jp).

II. BASIC PRINCIPLE OF THE PROPOSED ALGORITHM

In the followings let us to give a closer insight into details of the proposed algorithm. First of all let us emphasize the main principle together with the related problems. Let us refer to the stereo image pair as left and right images.

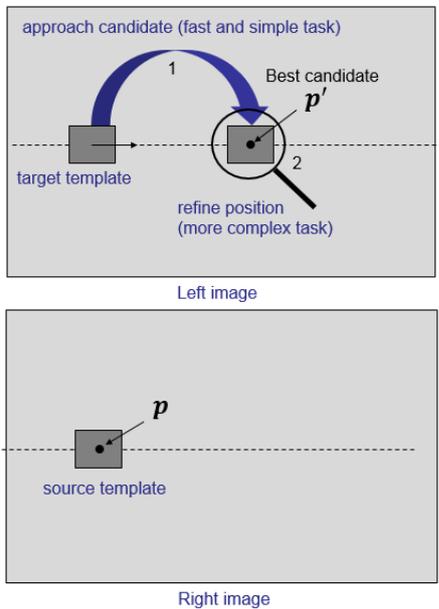


Fig. 1 Illustration of the two step based matching

As well known, when using local matching techniques, usually during the search for the best candidate point the neighborhood of many pixels along a scan line has to be examined by using some kind of similarity metric. In some specific cases the neighborhood of several hundred of templates has to be matched to a source template in the left image (depending on the baseline length, target's distance, etc.). Thus the complexity as well as the accuracy issues of the chosen metric are crucial in the matching process. On one hand simple methods are preferred because of their computational simplicity, but on the other hand in most of the cases their robustness to factors such as noise, pattern characteristics, illumination, occlusion, etc. is relatively weak compared to their counterpart methods, which are usually more robust to the mentioned factors, but are computationally more expensive. The proposed approach is based on the combination of simpler and more complex methods.

The algorithm can be divided into two main steps: In the first step the corresponding candidate's location is approximately determined by using a simpler matching approach (executed many times per search), followed by the refinement of the found candidate's position by a more robust and efficient technique (executed once per search). During the search for the candidate it is crucial to consider the influence of noise on the result.

After the corresponding points have been identified the 3D coordinates are determined by triangulation. The cameras were calibrated by the method proposed in [15]. Based on the

obtained intrinsic and extrinsic parameters all rays starting from the camera center and going through the given pixel are determined in advance for both cameras. Since the location of the corresponding point is going to be determined with sub-pixel precision, the corresponding camera rays should be estimated accordingly.

III. LOCATING THE CANDIDATE

In this section let us briefly summarize the consequences obtained through the analysis of the response of four window based similarity metrics, namely the sum of absolute differences (SAD), sum of squared differences (SSD), normalized sum of squared differences (NSSD) and the histogram of oriented gradients (HOG)[12].

According to our experiments – depending on the window size – the responses of the above metrics differ in general however in most of the cases the peak near the best candidate can be easily identified in their response. Depending on the texture it may happen that the response will contain more than one peak having near the same value (especially in case of repetitive patterns). In order to reduce such cases it is reasonable to consider metrics with high sensitivity on textural changes. At the same time it is also important to take into account their computational costs, since in case of each pixel these metrics should be evaluated more than once (depending on the disparity) in order to approach the location of the best corresponding candidate pixel. According to our experiments the NSSD and HOG reflect the highest sensitivity on textural changes, their response offer the most easily identifiable peak near the candidate pixel. Since the HOG is based on edge orientation histograms it is not advantageous for matching smoothly textured patches in the image. From computational complexity aspects it is the most expensive among all the above mentioned metrics. On the other hand the NSSD can offer useful response in either case, i.e. smooth textures as well as patches containing sharp edges. Hence in the upcoming sections let us assume that the location of the best candidate is estimated by NSSD. The most suitable window size may vary depending on the resolution and texture level of images, however the size of $32n \times 32n$ pixels, where $n = 1, 2, \dots$ is promising. In case of camera resolution 1280×960 pixels the most adequate window size for textures like skin or natural objects is 64×64 pixels (see Section V).

IV. SUB-PIXEL ACCURACY REFINEMENT OF THE LOCATED CANDIDATE

Let us assume that our $M \times N$ sized stereo images are related by $I_R(x, y) = I_L(x + t_x, y + t_y)$, where $\mathbf{t} = (t_x, t_y)$ stand for the displacement of two images. Our task is to determine t_x and t_y with sub-pixel accuracy. Let us note, that in this step the rough estimate of the corresponding point's location is already known (from the previous step). To refine the estimated location of the candidate, correlation techniques are utilized. Instead of estimating the correlation in spatial domain, it is much more efficient and computationally less expensive to get the translation vector through the normalized cross power spectrum (NCPS) of I_L and I_R .

A. Phase Correlation

Based on the convolution (how the convolution in the spatial domain relates to the multiplication in the frequency domain) and shift theorem (shift in the spatial domain results a linear phase difference in the frequency domain) of Fourier theory \mathbf{t} can be estimated as follows: Let U_R and U_L stand for the Fourier transforms (FT) of I_R and I_L respectively, i.e. $U_R(u, v) = \mathcal{F}\{I_R(x, y)\}(u, v)$ and $U_L(u, v) = \mathcal{F}\{I_L(x, y)\}(u, v)$. Based on the shift property of FT U_R and U_L are related as follows:

$$U_R(u, v) = U_L(u, v)e^{-j2\pi(\frac{ut_x}{M} + \frac{vt_y}{N})} \quad (1)$$

The NCPS of I_L and I_R can be expressed as follows:

$$\begin{aligned} \mathcal{S}(u, v) &= \frac{U_L(u, v)U_R(u, v)^*}{|U_L(u, v)U_R(u, v)^*|} = \\ &= \frac{U_L(u, v)U_L(u, v)^* e^{-j2\pi(\frac{ut_x}{M} + \frac{vt_y}{N})}}{|U_L(u, v)U_L(u, v)^*|} = e^{-j2\pi(\frac{ut_x}{M} + \frac{vt_y}{N})} \end{aligned} \quad (2)$$

The inverse FT of the above complex exponential in continuous case yields the shifted Dirac delta impulse $\delta(x - t_x, y - t_y)$. In order to reduce the noise, the frequencies are weighted by a Gaussian weighting function centered at the DC component. In this case $S(u, v)$ can be expressed as follows:

$$S(u, v) = G(u, v)e^{-j2\pi(\frac{ut_x}{M} + \frac{vt_y}{N})}, \quad (3)$$

where $G(u, v) = e^{-\left(\frac{u^2}{2\sigma_u^2} + \frac{v^2}{2\sigma_v^2}\right)}$. Based on the convolution theorem of FT, the inverse FT of complex exponential multiplied by a Gaussian is equivalent to the convolution of the Dirac delta with a Gaussian.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(p, q)\delta(x-t_x-p, y-t_y-q)dpdq = g(x-t_x, y-t_y),$$

and $g(x, y)$ stands for the inverse FT of $G(u, v)$, which is also a Gaussian function. Based on these considerations it is clear that in continuous case the model for $s(x, y)$ (the inverse FT of $S(u, v)$) is a Gaussian function (with proper parameters) centered at (t_x, t_y) . Since we are dealing with discrete images, let us examine the above relations for discrete case and find the best model to approximate $s(x, y)$. Let us discretize $S(u, v)$ over a two dimensional rectangular grid, having equidistant grid nodes. Let us denote the matrix element corresponding to grid point with coordinates u, v as $\mathcal{S}[u, v]$. The discrete inverse Fourier transform of $\mathcal{S}[u, v]$ can be expressed as follows:

$$s[x, y] = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \mathcal{S}[u, v] e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})}. \quad (4)$$

After substitution it can

$$s[x, y] = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} e^{-j2\pi(\frac{ut_x}{M} + \frac{vt_y}{N})} e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})}. \quad (5)$$

Let us express the above sum as the multiplication of two geometric series, i.e.

$$s[x, y] = \sum_{u=0}^{M-1} \left(e^{j\frac{2\pi}{M}(x-t_x)} \right)^u \sum_{v=0}^{N-1} \left(e^{j\frac{2\pi}{N}(y-t_y)} \right)^v.$$

After applying the rules for geometric series followed by simple rearrangements we get

$$s[x, y] = e^{j\pi(x-t_x)(1-\frac{1}{M})} \frac{\sin(\pi(x-t_x)) \sin(\pi(y-t_y))}{\sin(\frac{\pi}{M}(x-t_x)) \sin(\frac{\pi}{N}(y-t_y))}.$$

and for the normalized magnitude we get [14]:

$$\text{mag}\{s[x, y]\} = \frac{1}{MN} \frac{\sin(\pi(x-t_x)) \sin(\pi(y-t_y))}{\sin(\frac{\pi}{M}(x-t_x)) \sin(\frac{\pi}{N}(y-t_y))}. \quad (6)$$

The plot of the above function can be seen in Fig.2. Under mild conditions the above function can be approximated by the 2D sinc function. All the above considerations are true only for the ideal case, i.e. when the input images are noiseless.

However in case of measurements, data are corrupted by noise and distortion, thus the obtained NCPS \mathbf{S} will not fully reflect the properties of $\mathcal{S}(u, v)$ and hence its inverse FT will yield a noisy peak, which makes it difficult to estimate $[t_x, t_y]$ with sub-pixel accuracy. Thus, it is crucial to eliminate the noise and to guarantee that the constraints of $\mathcal{S}(u, v)$ are met.

Let us take a closer look at the properties of $\mathcal{S}(u, v)$. Based on (2) it is clear that \mathbf{S} can be expressed as the outer product of two complex valued vectors as follows (introduced in [13]):

$$\mathbf{S} = \mathbf{p}\mathbf{q}^T,$$

where $\mathbf{p}[u] = e^{-j2\pi\frac{ut_x}{M}}$ and $\mathbf{q}[v] = e^{j2\pi\frac{vt_y}{N}}$. Based on (2) in ideal case (i.e. noise free case) \mathbf{S} is a rank-1 matrix. This approach can be applied efficiently to estimate t_x and t_y directly in the frequency domain [13]. Usually images are corrupted by noise, thus the obtained NCPS will not fulfill the rank-1 constraint. Hence we are searching for the best rank-1 approximation $\tilde{\mathbf{S}}$ of \mathbf{S} . The best rank-1 approximation of \mathbf{S} giving the minimal Frobenius norm $\|\mathbf{S} - \tilde{\mathbf{S}}\|_F$ can be obtained by singular value decomposition (SVD) as follows:

$$\tilde{\mathbf{S}} = \sigma_1 \mathbf{w}_{1,1} \overline{\mathbf{w}}_{2,1}^T, \quad (7)$$

where $\mathbf{w}_{1,1}$ and $\mathbf{w}_{2,1}$ stand for the left and right singular vectors of \mathbf{S} corresponding to its largest singular value denoted

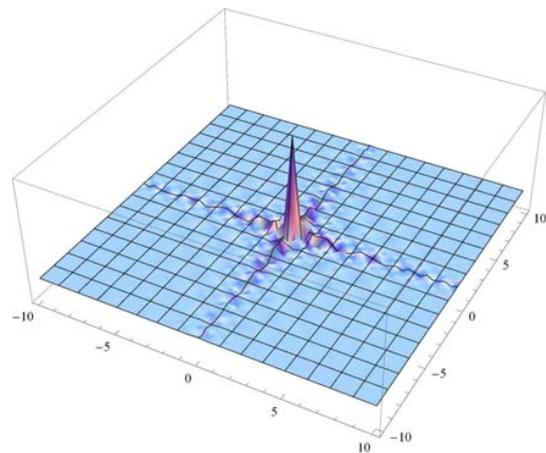


Fig. 2 The plot of $s[x, y]$. The discrete Fourier transform of a complex exponential

by σ_1 , respectively. By using these notations \mathbf{S} can be expressed similarly as follows:

$$\mathbf{S} = \sum_{i=1}^M \sum_{j=1}^N \sigma_{i,j} \mathbf{w}_{1,i} \overline{\mathbf{w}}_{2,j}^T. \quad (8)$$

In Fig. 4 the phase of the noisy NCPS and its rank-1 approximation can be followed. Although the rank-1 approximation clearly reflects the phase information of \mathbf{S} , it is still corrupted by noise in the high frequency region of the spectrum. Thus it is adequate to suppress the influence of these higher frequency components. For this purpose weighting by Gaussian function yields promising results (see Section V).

Weighting the frequencies by a Gaussian \mathbf{G} having the form of

$$\mathbf{G}[u, v] = e^{-\frac{\sigma^2(u^2+v^2)}{2}} \quad (9)$$

corresponds – as described for the continuous case – to the FT of

$$s[x, y] * \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x^2+y^2)}{\sigma^2}}, \quad (10)$$

where * denotes convolution. It is crucial to properly choose the variance σ^2 of the Gaussian. In case of such a suppression of higher frequencies the shape of the surface obtained by inverse FFT fits the exact model (6) more accurately (see Fig. 3).

However from computational complexity aspects applying SVD has strong negative impact on the processing time, since it has to run once for each image pixel followed by the inverse Fourier transformation of $\tilde{\mathbf{S}}$. Thus in case of using \mathbf{S} instead of $\tilde{\mathbf{S}}$ and applying Gaussian weighting the Gaussian model is suitable, however at the expense of accuracy. In order to obtain the relative shift $[t_x, t_y]$ between \mathbf{I}_L and \mathbf{I}_R , the following error function has to be minimized:

$$\min_{t_x, t_y, A} \left\{ \sum_{x,y} (\mathbf{C}[x, y] - \text{Model}(x, y, t_x, t_y, A)) \right\}, \quad (11)$$

where matrix \mathbf{C} denotes the inverse FT of $\tilde{\mathbf{S}}$ and Model represents (depending on the applied filter) $As[x, y]$ or $Ae^{-\sigma^2((x-t_x)^2+(y-t_y)^2)}$

In order to reduce the spectral leakage caused by the edge discontinuities (i.e. if the pixel values at the opposite edges of the image are different) the image is multiplied by the Hann window, defined as follows:

$$W(x, y) = 0.25 \left(1 - \cos\left(\frac{2\pi x}{M}\right) \right) \left(1 - \cos\left(\frac{2\pi y}{N}\right) \right). \quad (12)$$

Algorithm 1 contains the pseudo-code of the proposed method.

V. MODEL EVALUATION AND EXPERIMENTAL RESULTS

In this section we would like to show the efficiency of the *Gaussian* and *sinc* model as well as the suitability of the proposed matching method for passive 3D measurement. Based on two images related by a known horizontal and vertical translation $\mathbf{t}_* = [t_{x_*}, t_{y_*}]$ the suitability of the models discussed in the previous section have been validated in case of

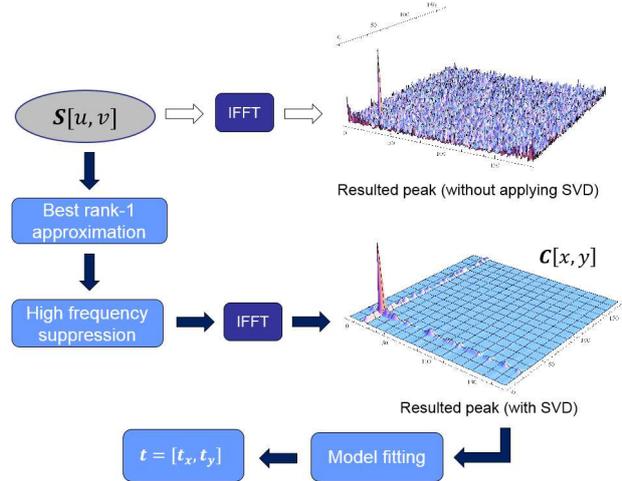


Fig. 3 Illustrating the flow of refinement.

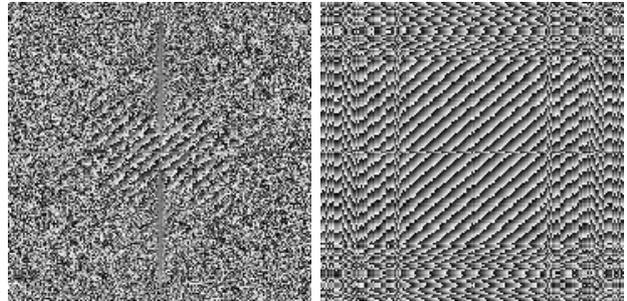


Fig. 4 The phase of \mathbf{S} (left) and its rank-one approximation $\tilde{\mathbf{S}}$ (right)

unfiltered as well as filtered NCPS. In Tables I and II the true and the estimated values of t_x and t_y as well as the residual mean square error (RMSE) can be followed.



Fig. 5 Two templates relatively shifted by $t_x = 11.5\text{pixels}$, $t_y = 11.5\text{pixels}$

In ideal case the location of the maximal value of \mathbf{S} corresponds to \mathbf{t}_* , however when dealing with noisy data this assumption will be not true, i.e we have to improve the location of the best candidate by model fitting. Let us choose the starting point $[t_{x_0}, t_{y_0}]$ for the optimization (11) to be equal the location of the maximal value of NCCS. Let us show how the models perform in such a \mathbf{t}_0 in case of filtered and non-filtered spectra, as well. As the data reflect, both models (the Gaussian as well as the $s(x,y)$) are suitable for practical use, however in case of the Gaussian model, the value of the σ parameter is crucial. If σ is too small, i.e. the Gaussian hump is narrower, we can easily get nipped in a local minimum. We examined the effect of a smaller and a greater σ , as well (see Table I

Algorithm 1 Matching

```

1: Let assume a rectified stereo image pair  $\mathbf{I}_L, \mathbf{I}_R$ 
2:  $[w, h] \leftarrow Template\_size$ 
3:  $\mathbf{p}_L = [x_L, y_L]$ 
4: for all pixels  $\mathbf{p}_R = [x_R, y_R]$  in  $\mathbf{I}_R$  do
5:    $\mathbf{p}_L \leftarrow \mathbf{p}_R$ 
6:    $\mathbf{T}_L(\mathbf{p}_L) \leftarrow$  rectangular template centered at  $\mathbf{p}_L$ 
7:    $\mathbf{T}_R(\mathbf{p}_R) \leftarrow$  rectangular template centered at  $\mathbf{p}_R$ 
8:    $\mathbf{T}_R \leftarrow \mathbf{W} \circ \mathbf{T}_R$ 
9:    $\mathbf{U}_R \leftarrow FFT(\mathbf{T}_R)$ 
10:   $d \leftarrow threshold$ 
11:   $N \leftarrow step\_size$ 
12:  repeat
13:     $\min_{\Delta x} \{NSSD(\mathbf{T}_R(x_R, y_R), \mathbf{T}_L(x_L + \Delta x, y_L))\}$ 
    subject to  $\Delta x \bmod N = 0; 0 < \Delta x < d$ 
14:     $x_L \leftarrow x_L + \Delta x - N$ 
15:     $d \leftarrow N$ 
16:     $N \leftarrow N - 1$ 
17:  until  $N = 0$ 
18:   $\mathbf{T}_L \leftarrow \mathbf{W} \circ \mathbf{T}_L(\mathbf{p}_L)$ 
19:   $\mathbf{U}_L \leftarrow FFT(\mathbf{T}_L)$ 
20:   $\mathbf{S} \leftarrow NCPS(\mathbf{U}_R, \mathbf{U}_L)$ 
21:   $\tilde{\mathbf{S}} \leftarrow$  best rank-1 approximation of  $\mathbf{S}$ .
22:   $\tilde{\mathbf{S}} \leftarrow \mathbf{G} \circ \tilde{\mathbf{S}}$  ( $\circ$  denotes the Hadamard product)
23:   $\mathbf{C} \leftarrow IFFT(\tilde{\mathbf{S}})$ 
24:   $\min_{t_x, t_y, A} \{ \sum_{x, y} (\mathbf{C}[x, y] - Model(x, y, t_x, t_y, A, \sigma)) \}$ 
25: end for

```

TABLE I
MODEL EFFICIENCY IN CASE OF ORIGINAL, RANK-1 AND SMOOTHED
NCPS

| NCPS | s(x,y) | | | Gaussian ($\sigma = 0.5$) | | |
|---------------------------------------|--------|-------|--------|-----------------------------|-------|--------|
| | t_x | t_y | RMSE | t_x | t_y | RMSE |
| $\tilde{\mathbf{S}}$ | 11.34 | 11.13 | 0.0009 | 11.37 | 11.12 | 0.0000 |
| $\mathbf{G} \circ \tilde{\mathbf{S}}$ | 11.12 | 11.38 | 0.0010 | 11.48 | 11.48 | 0.0000 |
| $\mathbf{G} \circ \mathbf{S}$ | 11.08 | 11.35 | 0.0584 | 11.00 | 11.01 | 0.0659 |
| \mathbf{S} | 11.51 | 11.07 | 0.9767 | 11.01 | 11.01 | 0.9948 |
| \mathbf{t}_* | 11.50 | 11.50 | | 11.50 | 11.50 | |
| \mathbf{t}_0 | 11.00 | 11.00 | | 11.00 | 11.00 | |

and II). If Gaussian model fitting is performed on the inverse FT of the best rank-1 approximation of NCCS, the results are less sensitive to the σ value. However in case of \mathbf{S} and $\mathbf{G} \circ \mathbf{S}$ the σ value is much more crucial. Regarding the model $s(x, y)$ we can say that it performs better on SVD filtered data, and it does not need any additional parameter to tune the model. The Gaussian model is promising for both the SVD as well as the Gaussian filtered case, however depending on the σ parameter.

From the fitting accuracy point of view the rank-1 approximation of the NCCS gives the best results, as its shape characteristics are close to the ideal model $s(x, y)$ (see Figs. 2 and 3). In order to show the performance and suitability of the method for passive 3D reconstruction, a human face has been measured (see Figs. 7-10). As the method does not require any pattern projection for the measurement it is suitable for measuring moving objects. Furthermore in contrast to the most of active scanners it is adequate for outdoor applications, as

TABLE II
MODEL EFFICIENCY IN CASE OF ORIGINAL, RANK-1 AND SMOOTHED
NCPS

| NCPS | s(x,y) | | | Gaussian ($\sigma = 1.0$) | | |
|---------------------------------------|--------|-------|--------|-----------------------------|-------|--------|
| | t_x | t_y | RMSE | t_x | t_y | RMSE |
| $\tilde{\mathbf{S}}$ | 11.34 | 11.13 | 0.0009 | 11.27 | 11.29 | 0.0001 |
| $\mathbf{G} \circ \tilde{\mathbf{S}}$ | 11.12 | 11.38 | 0.0010 | 11.35 | 11.41 | 0.0002 |
| $\mathbf{G} \circ \mathbf{S}$ | 11.08 | 11.35 | 0.0584 | 11.25 | 11.32 | 0.0632 |
| \mathbf{S} | 11.51 | 11.07 | 0.9767 | 11.37 | 11.28 | 0.9893 |
| \mathbf{t}_* | 11.50 | 11.50 | | 11.50 | 11.50 | |
| \mathbf{t}_0 | 11.00 | 11.00 | | 11.00 | 11.00 | |

well. The limitation of the method is, that the surface should contain at least a small amount of texture in order to be able to find the best candidate by NSSD and phase correlation in contrast to the active measurement methods.



Fig. 6 The stereo image pair used for reconstruction



Fig. 7 Reconstructed face surface by the proposed method (view-1)



Fig. 8 Reconstructed face surface by the proposed method (view-2)



Fig. 9 Reconstructed face surface by the proposed method (view-3)



Fig. 10 Reconstructed face surface by the proposed method (view-4)

VI. CONCLUSION

In this paper a passive stereo vision based 3D measurement approach was proposed. The method is based on approaching the best candidate by a simple NSSD based method followed by a phase correlation based refinement step which locates the best candidate with sub-pixel accuracy. As the results reflect, the method is suitable to reconstruct low textured areas, as well. The rank-1 approximation of the noisy NCPS makes the method more robust to noise. Although the computational complexity of the method is high, it can efficiently be parallelized.

ACKNOWLEDGMENT

This research was supported by the JSPS Invitation Fellowship Program for Research in Japan FY 2014 and by the by JSPS KAKENHI Grant Number 25289123.

REFERENCES

- [1] M. Young, E. Beeson, J. Davis, S. Rusinkiewicz and R. Ramamoorthi, "Viewpoint-Coded Structured Light", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 18, June 2007

- [2] E. Lilenblum, B. Michaelis, "Optical 3D Surface Reconstruction by a Multi-Period Phase Shift Method, Journal of Computers, Vol. 2, No. 2, pp. 73-83, 2007
- [3] Yan Cui, Schuon, S., Chan, D., Thrun, S., Theobalt, C., "3D Shape Scanning with a Time-of-Flight Camera," 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1173-1180, 13-18 June 2010
- [4] N. Karpinsky, S. Zhang, "High-Resolution, Real-Time 3D Imaging with Fringe Analysis," Journal of Real-Time Image Processing, Springer-Verlag, Vol. 7, Issue 1, pp. 55-66, 2010
- [5] Raj Kumar Gupta, Siu-Yeung Cho, "Window-Based Approach for Fast Stereo Correspondence," IET Computer Vision, Vol. 7, Issue 2, ISSN 1751-9632, pp. 123-134, April 2013.
- [6] Adhyapak Satyajit Anil, Kehtarnavaz Nasser, Nadin Mihai, "Stereo matching via selective multiple windows," Journal of Electronic Imaging Vol. 16, No. 1, pp. 013012-1-013012-14, 2007.
- [7] Boykov, Y.; Veksler, O.; Zabih, R., "Fast approximate energy minimization via graph cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 11, pp. 1222-1239, Nov. 2001.
- [8] M. Bleyer, M. Gelautz, "Graph-Cut-Based Stereo Matching Using Image Segmentation with Symmetrical Treatment of Occlusions, Signal Processing: Image Communication, Vol. 22, Issue 2, ISSN 0923-5965, pp. 127-143, Feb. 2007.
- [9] Liu, Yonghuai, "A Mean Field Annealing Approach to Accurate Free Form Shape Matching," Journal of Pattern Recognition, Vol. 40, Issue 9, ISSN: 0031-3203, pp. 2418-2436, 2007.
- [10] Jian Sun; Nan-Ning Zheng; Heung-Yeung Shum, "Stereo matching using belief propagation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 7, pp. 787-800, July 2003.
- [11] Sawires, E.F.; Hamdy, A.M.; Amer, F.Z.; Bakr, E. M., "Disparity map using suboptimal cost with dynamic programming," IEEE International Symposium on Signal Processing and Information Technology, pp. 209-214, 15-18 Dec. 2010.
- [12] Kachouane, M.; Sahki, S.; Lakrouf, M.; Ouadah, N., "HOG based fast human detection," 24th International Conference on Microelectronics (ICM), pp. 1-4, 16-20 Dec. 2012.
- [13] Hoge, W.S., "A Subspace Identification Extension to the Phase Correlation Method [MRI application]," IEEE Transactions on Medical Imaging, Vol. 22, No. 2, pp. 277-280, Feb. 2003.
- [14] Muquit, M. A., Shibahara, T., Aoki, T., "A High-Accuracy Passive 3D Measurement System Using Phase-Based Image Matching," IEICE Transactions, Vol. E89, No. 3, pp. 686-697, 2006.
- [15] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.