New Insights into the Calibration of ToF-Sensors

Marvin Lindner
Inst. for Vision and Graphics
University of Siegen, Germany
marvin.lindner@uni-siegen.de

Andreas Kolb
Inst. for Vision and Graphics
University of Siegen, Germany
andreas.kolb@uni-siegen.de

Thorsten Ringbeck
PMD Technologies
Siegen, Germany
t.ringbeck@pmdtec.com

Abstract

Time-of-Flight (ToF) sensors have become an alternative to conventional distance sensing techniques like laser scanners or image based stereo. ToF sensors provide full range distance information at high frame-rates and thus have a significant impact onto current research in areas like online object recognition, collision prevention or scene reconstruction.

However, ToF cameras like the Photonic Mixer Device (PMD) still exhibit a number of challenges regarding static and dynamic effects, e.g. systematic distance errors and motion artefacts, respectively. Sensor calibration techniques reducing static system errors have been proposed and show promising results. However, current calibration techniques in general need a large set of reference data in order to determine the corresponding parameters for the calibration model.

This paper introduces a new calibration approach which combines different demodulation techniques for the ToF-camera's reference signal. Examples show, that the resulting combined demodulation technique yields improved distance values based on only two required reference data sets.

Furthermore, we discuss a specific effect of the so-called pre-adjustment step suitable for high-order calibration techniques based on b-splines.

1. Introduction

In automatization areas like robotics or automotive engineering, the reconstruction of objects and scenes is a necessary fundamental with respect to computer vision. Information obtained from digitized scenes represent important input data for position determination, online object recognition, or collision prevention.

During the last years, a compact and low-priced alternative to common laser scanners and stereo-vision setups has gained in popularity. Based on the time-of-flight principle, ToF-cameras like the Photonic Mixer Device (PMD) are capable to estimate full-range distance information in real time, i.e. with up to 20 fps, by illuminating the scene with modulated infrared light and determining the phase-shift between the reference signal and the reflected light (see Sec. 2).

Unfortunately, several error sources necessitate a proper calibration of such devices in order to get accurate distance information. For instance, the measured distance of a PMD is affected by a systematic error [2, 5], the integration time and the amount of incident active light [6].

The contribution of this paper is an alternative demodulation scheme for ToF sensors that are based on the sampling of an internal auto-correlation function (ACF). We discuss a demodulation approach based on the assumption of a rectangular-shaped reference signal. In addition, we combine this approach with the frequently used sinusoidal reference signal (see e.g. [4]) resulting in an efficient calibration technique that requires two reference distance measurements only. Furthermore, we discuss the relation between a standard fixed-pattern noise correction and the so-called *pre-adjustment* proposed by Lindner and Kolb[5] in the context of high-order distance error correction based on b-splines.

After a short introduction to the technological foundation in Sec. 2, an overview of the current calibration models and reference data acquisition approaches is given in Sec. 3. An alternative demodulation of the auto-correlation function for box signals and its application for an alternative distance calibration model is considered in Sec. 4. Finally, the results of the alternative distance calibration model and some aspects for the high-order b-spline calibration are discussed in Sec.5.

2. Technological Foundation

The main component of common phase-based ToF-cameras is a special sensors consisting of so-called smart pixels [3, 4, 11], each correlating the outgoing reference signal s with the incoming optical signal, i.e. the reflected

signal, r yielding

$$c(\tau) = s \otimes r = \lim_{T\to\infty} \int_{-T/2}^{T/2} s(t) \cdot r(t+\tau)\, dt.$$

In many approaches a sinusoidal signals is assumed,

$$s(t) = \cos(\omega t), \qquad r(t) = k + a\cos(\omega t + \phi) \tag{1}$$

where ω is the angular frequency, a is the amplitude of the incident optical signal and ϕ is the phase offset relating to the object distance, some trigonometric calculus yields $c(\tau) = \frac{a}{2}\cos(\omega\tau + \phi)$.

By sampling the correlation function four times $A_i = c(i \cdot \pi/2\omega)$, a ToF-camera is capable to determine a pixel's phase shift ϕ_{sin}, the correlation amplitude a and the incident light intensity b by

$$\phi_{\text{sin}} = \arctan\left(\frac{A_3 - A_1}{A_0 - A_2}\right), b = \tfrac{1}{4}\sum_{i=0}^{3} A_i, \tag{2}$$
$$a = \tfrac{1}{2}\sqrt{(A_3 - A_1)^2 + (A_0 - A_2)^2}$$

The distance to the according object region is $d = \frac{c}{4\pi\omega}\phi$, where $c \approx 3 \cdot 10^8 \frac{m}{s}$ is the speed of light.

Note, that theoretically three samples are sufficient, but due to stability considerations, four samples are commonly used.

3. Prior Work

The assumption of a sinusoidal correlation function as in Sec. 2 is not met for existing sensors. Due to hardware and cost limitations, it is practically not feasible to generate a perfect sinusoidal reference signal. Analyzing the real reference signal of a PMD camera, it arises that the optical signal shape is rather far from the theoretical assumed sinusoidal shape. [8]. The result is a systematic distance error as shown in Fig. 1.

At present, two major directions to handle the systematic error of ToF-cameras exist. On the one hand, a more accurate representation of the ACF and the corresponding reconstruction is discussed (see Sec. 3.1), on the other hand methods have been proposed to correct the systematic distance error by phenomenological calibration models (see Sec. 3.2).

3.1. Higher Order Demodulation

Assuming the correlation to be an ideal convolution as depicted in Eq. 1, the correlation function for nonharmonical signals typically consists of higher Fourier modes.

Therefore, Langer [4] and Rapp [8] discuss an enhanced representation of the ACF modeled by a finite sum of superimposed cosine waves

$$c(\tau) = \sum_{k=0}^{l} c_k \cos(k(\omega\tau + \phi) + \theta_k).$$

A least square optimization over $N \geq 2l+1$ samples of the ACF leads to following phase demodulation schema:

$$k\phi + \theta_k = \arg\left(\sum_{n=0}^{N-1} A_n e^{-2\pi i k \frac{n}{N}}\right)$$

where $A_n = c(\frac{2\pi}{\omega} \cdot \frac{n}{N})$. Finally, the distance related phase-shift ϕ can be obtained by using a look-up table for the fixed offsets θ_k of the additional modes.

In practice, extending the demodulation theme for non-harmonically signals is impracticable as the number of required sample images as well as the calculation effort for the demodulation would be too large. Especially the higher number of samples leads to further interferences in acquiring dynamic scenes.

3.2. Correction of Distance Errors

3.2.1 Modelling the Distance Error

Simple calibration models for phase-based ToF-cameras try to model the systematic error by linear or polynomial functions [1, 10] or as fixed-pattern noise. For some special cases, e.g. a rather small range of interest, this approach might be a suitable, but in general they restrict the working range of the sensor or the accuracy (see also Fig. 1).

More accurate models use look-up tables [2] or higher order functions e.g. modeled by b-splines [5] in order to express the systematic error more precisely. These models provide a much better error compensation, but result in an increased calibration effort. Especially the required amount reference data and calibration data is required.

The b-spline approach introduced by Lindner and Kolb [5] first transforms the measured and the reference data into cartesian coordinates resulting in measured distances $m_k(x, y)$ at pixel (x, y) with reference distance

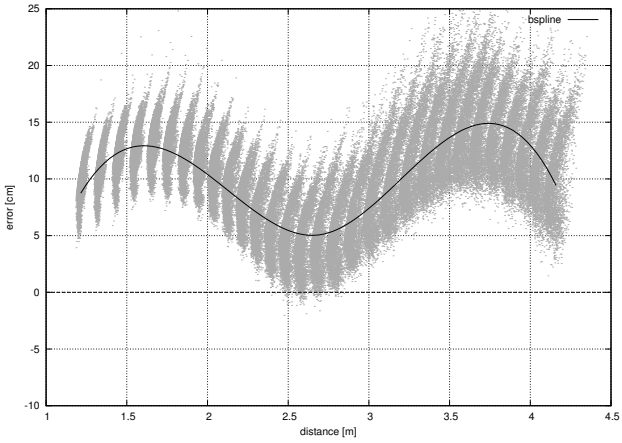

Figure 1. Demodulation error in [cm] over the interval 1-4 m with a b-spline fitting (solid line).

(ground truth) $d_k(x, y)$ for the k-th reference distance. The method mainly consists of three steps:

1. Determine a global b-spline $d = b_{\text{glob}}(m)$ subject to the least-squares minimization of $\sum_{(x,y),k} \left(b(m_k(x, y)) - d_k(x, y)\right)^2$

2. A per-pixel pre-adjustment using a linear function $l^{\text{pre}}_{x,y}$ per-pixel (x, y) subject to the least squares minimization of $\sum_k \left(m_k(x, y) - m^{\text{avg}}_k\right)^2$, where m^{avg}_k is the average over all distance measurements for the k-th reference distance. The resulting correction after this step is

$$\tilde{m}(x, y) = b_{\text{glob}}(m(x, y) - l^{\text{pre}}_{x,y}(m(x, y))).$$

3. Post-adjustment further reduces the remaining distance error using a second linear function per pixel $l^{\text{post}}_{x,y}$. The final correction after this step is

$$\tilde{\tilde{m}}(x, y) = \tilde{m}(x, y) - l^{\text{post}}_{x,y}(\tilde{m}(x, y))).$$

The second step accounts for the fact, that the b-spline corrects the *average* systematic distance error, but individual pixel offsets lead to a evaluation of the b-spline at a wrong distance, respectively phase, and thus to a wrong distance correction.

Note, that both linear per-pixel correction steps can be used optionally.

3.2.2 Acquisition of Reference Data

All calibration approaches need a rather dense set of reference data, which makes the calibration process fairly complex with increasing accuracy.

The most obvious and simple approach for this task covers the utilization of special track lines, which can be used to automatically address precise distances to e.g. a plane wall [2, 5].

Other approaches try to avoid the need for special equipment by using vision based algorithms [6] optional improved by regression methods in combination with stereo-vision like setups [1]. Here, position and orientation of the reference plane are determined by the camera's extrinsic parameters in respect to a special marker. The marker is detected in the PMD's intensity or an optional second high-resolution camera image and it pose is computed using standard CV methods [7].

For low-resolution ToF-cameras like the 3k-PMD the utilization of regression methods turned out to be recommendable. Here, the extrinsic parameters are additionally improved by an iterative optimization approach. In each step a synthetic view for the given set of parameters is generated which leds to a successive parameter adjustment [9].

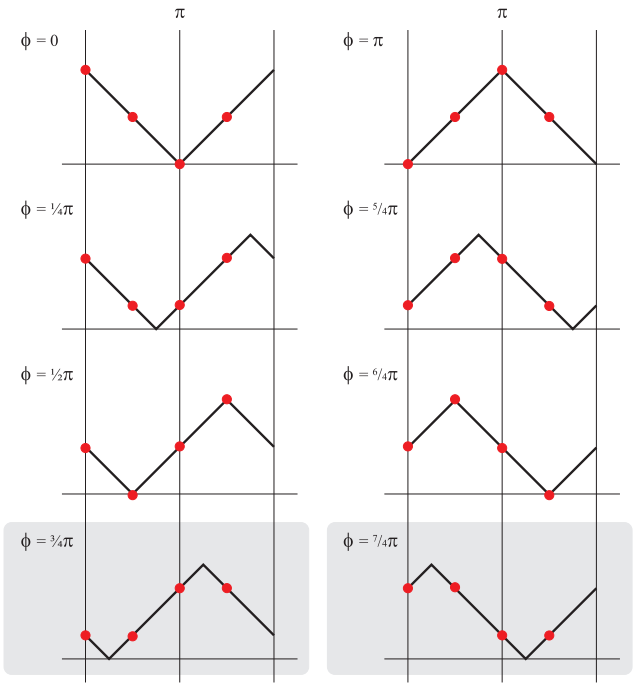

Figure 2. Sampling values for a triangular correlation function and their corresponding phase offset. Special care must be taken for the gray shaded cases.

4. Alternative Calibration Approach

The improvement of distance accuracy by either enhanced demodulation schemes or heuristic calibration models usually implies extra effort by means of hardware modification or reference data acquisition. For this reason, an alternative approach would be desirable, which uses the standard four ACF samples and, at the same time, less reference data compared to existing models.

4.1. Demodulation of a Triangular ACF

For Photonic Mixer Devices, measurements reveal that the reference signal exhibits a mixture between a rectangular and a sinusoidal shape [8], which leads to an alternative demodulation approach.

Assuming an ideal rectangular signal, the correlation function c is triangular with its valley points displaced by the phase shift ϕ_{tri} (see Fig. 2).

By fitting two intersecting lines

$$l_{1,2}(\theta) = m_{1,2} \cdot (\theta - \phi'_{\text{tri}}) + t$$

with $m_1 = -m_2$ through the sample points A_i, the phase offset ϕ_{tri} can be obtained by

$$\phi_{\text{tri}} = \underbrace{\left(\frac{\pi}{2} \cdot \frac{A_3 - A_1}{A_0 - A_2 + A_3 - A_1}\right)}_{\pi - \phi'_{\text{tri}}} + \begin{cases} \pi & A_0 < A_1 \wedge A_2 > A_3 \\ 0 & \text{else} \end{cases}$$

The amplitude a can be calculated by

$$a = \frac{1}{4}\sum_{i=0}^{3} A_i - t$$

However, special care most be taken for phase shifts where the valley points are located between the first two and the last two sample points (gray shaded cases in Fig. 2). In this case, the last sample point must be moved to the front in order to establish the right fitting situation. This means that A_i becomes $A_{(i+1) \mod 4}$ whereas the intersection point t is shifted by an additional amount of $\pi/2$.

4.2. Combined Demodulation

Applying the new demodulation approach, the distance error unfortunately can not be reduced compared to the sinusoidal case (see Fig. 3). However, the new sampling approach can still be used to attenuate the distance error. As the error trend is inverse to the systematic error for sinusoidal demodulation, a linear combination

$$\phi = a \cdot \phi_{\text{sin}} + b \cdot \phi_{\text{tri}} + c \tag{3}$$

seems to be suitable to compute a new phase offset with higher accuracy than the one provided by the individual demodulation schemes.

Analog to existing heuristic calibration modules, the optimal linear combination a, b and c can be find by least square optimization in respect to known reference data.

To keep the number calibration parameter as small as possible, we decided to let c be a constant per-pixel offset comparable to *fixed pattern noise*, whereas a and b correspond to global calibration parameters.

5. Results and Discussion

5.1. Combined Demodulation

Our measurements have shown that the combined demodulation approach actually can not keep up with the b-spline approach presented in [5], but in contrast is fairly independent to the number of reference images (see Fig. 4).

Combined demodulation is therefore very effective in means of the required reference data, i.e. two reference images are already adequate to archive good results. An accurate b-spline in contrast needs about 12 - 16 well chosen reference images to avoid undersampling. As a result, an acceptable distance adjustment can be archived for a minimum number reference data using the proposed combined demodulation in Eq. 3.

Compared to a constant or linear adjustment of the original demodulation scheme, i.e. $a = 1$ and $b = 0$, the combined demodulation model gives clearly the best results as it is the only technique out of these three which is able to cope with the systematic error (see Fig. 5).

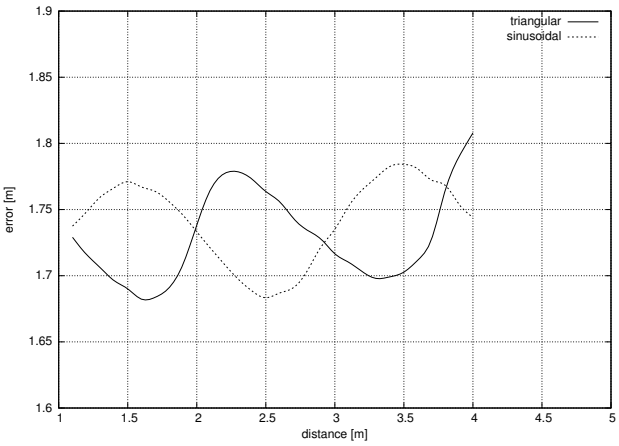

Figure 3. Mean distance error for the original and the new triangle-based demodulation approach.

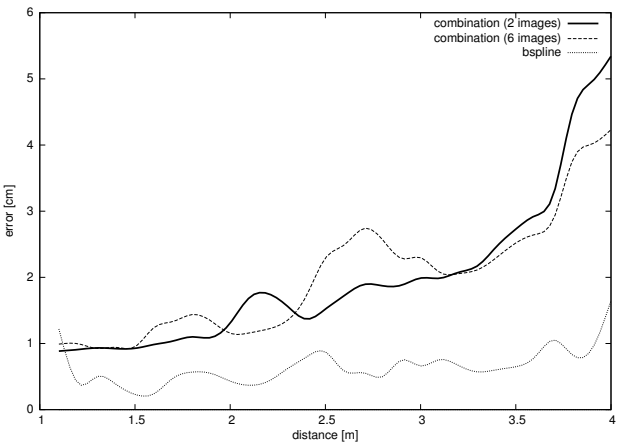

Figure 4. Mean distance error for the combined demodulation approach using two (solid) or six reference images (dashed line) compared to the optimal b-spline results.

Improved results can be archived by high-order combinations of ϕ_{sin} and ϕ_{tri} like

$$\phi = \sum_{i,j=0}^{2} a_{ij} \cdot \phi_{\text{sin}}^{i} \phi_{\text{tri}}^{j}.$$

Unfortunately, this way the number of parameters and therefore the number of necessary reference images increases, which in result brings no real advantages compared to e.g. the b-spline approach [5].

5.2. B-Spline Calibration Scheme

Evaluating the different calibration methods and comparing them to the combined demodulation scheme proposed in Sec. 4, some interesting results regarding the b-spline technique introduced by Lindner and Kolb [5] (see also Sec. 3.2.1) could be derived.

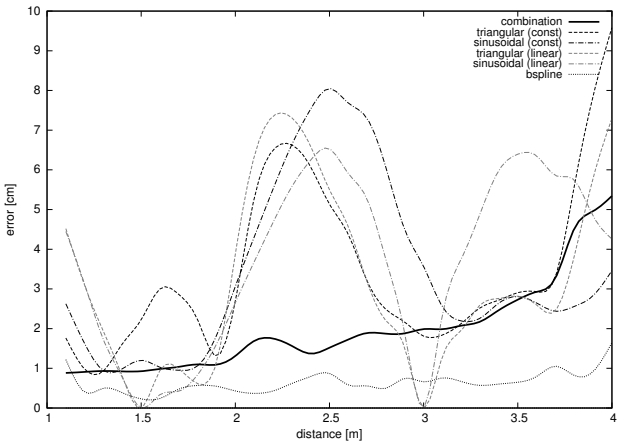

Figure 5. Mean distance error for the combined demodulation approach (solid line) compared to constant and linear per pixel adjustment (dashed lines). The b-spline results are added for completeness.

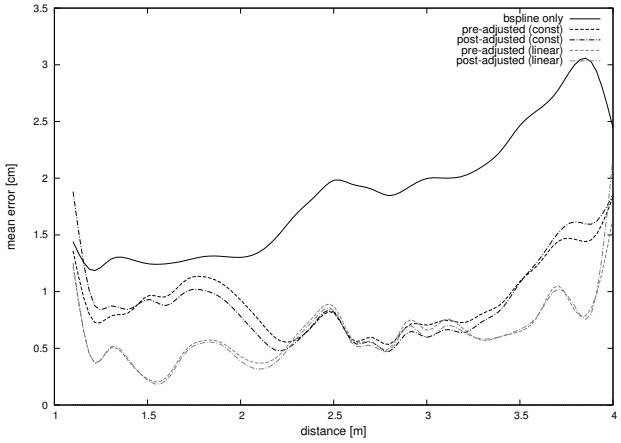

Figure 6. Mean error after the b-spline correction, including only per-pixel pre-adjustment and including only per-pixel post-adjusted.

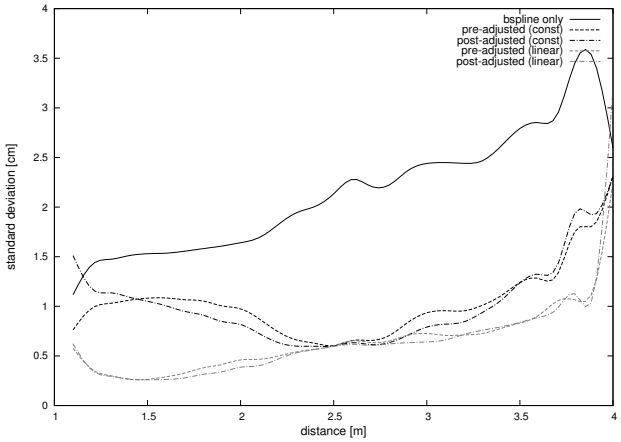

Figure 7. Standard error deviation after the b-spline correction, including only per-pixel pre-adjustment and including only per-pixel post-adjusted.

Concerning the per-pixel approach in [5], there have been some discussions and two strong conjectures:

1. linear correction does not significantly improve the results compared to a constant per-pixel correction (fixed pattern noise)

2. pre-adjustment is basically the same as post-adjustment, i.e. a linear correction applied to the remaining distance error

Our test invalidated the first conjecture and validated the second. Even though the linear coefficients for the per-pixel correction are in general very small (about 10^{-3} at average), the linear correction exhibits a significant improval over the simple usage of constant per-pixel correction approaches (see Fig. 6). This is also true, if we consider the standard deviation for the distance error (see Fig. 7).

The main argument against the second conjecture is the fact, that the per-pixel offsets are applied to a non-linear correction function. However, the results reveal that the application of the pre-adjustment or alternatively the post-adjustment yield very similar results. Obviously, the non-linearity of the distance correction, i.e. the b-spline is small enough in the range of the per-pixel displacements, resulting in an overall nearly linear per-pixel correction effect.

Concerning the effects of applying both, the pre- and the post-adjustment, we found, that there is only a little improvement. This has already been reported in [5].

6. Conclusion

We discussed a demodulation approach for an alternative triangle-shaped correlation function for ToF-sensors. Combining this approach with the known demodulation for sinusoidal signals, we found a way to improve the distance accuracy on the basis of two reference images only.

Furthermore, we discussed some aspects of the high-order b-spline calibration method proposed by Lindner and Kolb [5]. As a result, we could show, that either pre- or post-adjusted results based on linear correction functions is a good compromise between quality on the one hand and and computatial and storage costs on the other hand.

Acknowledgement

Will be added in the final version.

References

[1] C. Beder and R. Koch. Calibration of focal length and 3D pose based on the reflectance and depth image of a planar object. *Int. J. on Intell. Systems Techn. and*

App., Issue on Dynamic 3D Imaging, 2008. accepted for publication. 2, 3

[2] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration for increased accuracy of the range imaging camera SwissRanger^{TM}. In *Image Engineering and Vision Metrology (IEVM)*, 2006. 1, 2, 3

[3] H. Kraft, J. Frey, T. Moeller, M. Albrecht, M. Grothof, B. Schink, H. Hess, and B. Buxbaum. 3D-camera of high 3D-frame rate, depth-resolution and background light elimination based on improved PMD (photonic mixer device)-technologies. In *OPTO*, 2004. 1

[4] R. Lange. *3D Time-Of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology*. PhD thesis, University of Siegen, 2000. 1, 2

[5] M. Lindner and A. Kolb. Lateral and depth calibration of PMD-distance sensors. In *Int. Symp. on Visual Computing (ISVC)*, volume 2, pages 524–533. Springer, LNCS, 2006. 1, 2, 3, 4, 5

[6] M. Lindner and A. Kolb. Calibration of the intensity-related distance error of the PMD ToF-camera. In *Proc. SPIE, Intelligent Robots and Computer Vision*, volume 6764, 2007. doi:10.1117/12.752808. 1, 3

[7] OpenCV, 2006. http://sourceforge.net/projects/opencvlibrary. 3

[8] H. Rapp. Experimental and theoretical investigation of correlating TOF-camera systems. Master's thesis, University of Heidelberg, Germany, 2007. 2, 3

[9] I. Schiller, C. Beder, and R. Koch. Calibration of a pmd camera using a planar calibration object together with a multi-camera setup. In *Proceedings of the XXI. ISPRS Congress, Bejing, China*, 2008. to appear. 3

[10] M. Stommel and K.-D. Kuhnert. Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4780–4785, October 9-15, 2006. 2

[11] Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, and T. Ringbeck. Smart pixel – photonic mixer device (PMD). In *Proc. Int. Conf. on Mechatron. & Machine Vision*, pages 259–264, 1998. 1