Proceedings of 16th Int. Workshop on Robotics in Alpe-Adria-Danube Region - RAAD 2007
Ljubljana, June 7-9, 2007

COMPARISON AND EVALUATION OF SIMILARITY MEASURES FOR VERGENCE ANGLE ESTIMATION[1]

Dimitrios Chrysostomou
Antonios Gasteratos

Laboratory of Robotics and Automation, Department of Production and Management Engineering,
Democritus University of Thrace,
GR-671 00 Xanthi, Greece

Abstract: This paper presents a comparison of various similarity measures developed for the real-time control of the vergence angle in an active vision robot head. The vergence angle can be estimated using several difference or correlation measures. These methods are studied comparatively for various image sizes. The Zero-mean Normalized Cross Correlation (ZNCC) measure proved to outperform the other methods. The results also show that we can sufficiently control the vergence mechanism, using images even 256 times smaller than the original (i.e. 40x30 pixels), in less than 1ms. We evaluated these results using by extracting the disparity maps of the stereo pair to test whether the disparity value is zeroed, when we reach the correct vergence angle. We also calculated the Mean Square Error (MSE) and the Normalized MSE (NMSE) of the correlation index between the sub-sampled and the initial images.

Keywords: Vergence Angle, Similarity Measures, Binocular Disparity.

1. INTRODUCTION

The control of vergence angle has a primary role in active vision stereo systems. This is the angle between the two optical axes of the cameras. The vergence angle is related to the distance between the cameras and the fixation point. Successful control of the vergence angle affects many attention related applications including tracking and gaze holding, as well as depth estimation and 3D scene reconstruction. Our application aims at the control of a robotic head mounted on a robot for bomb disposal, which operates mainly outdoors (Arz *et al.*, 2005). Several algorithms have been proposed, in order to calculate and control the vergence angle in various stereo robotic heads. These are based on the estimation of binocular disparity (Hansen and Sommer, 1997; Hee-Jong *et al.*, 2000) or on correlation (Bernardino and Santos-Victor, 1996; Manzotti *et al*., 2001), whilst methods inspired by the human visual system appeared in the recent literature (Sang-Bok *et al.*, 2004; Tsang and Shi, 2006).

Binocular disparity measures the difference of a scene as projected on the two eyes. A great variety of methods for the binocular disparity calculation of a stereo pair of images exist. These are based on correlation measures (Niitsuma and Maruyama, 2004; Kotoulas *et al*., 2005), phase differences (Diaz *et al.,* 2007) and dynamic programming (Sun, 2003; Forstmann *et al.,* 2004) to mention some. However, in order to use the above algorithms, precise and accurate calibration of the stereo pair is essential. It is common truth that the stereo camera calibration is a difficult and laborious process. Also, in outdoor environments there are hardly any means to correct the precision of the calibration, in case that is needed.

Thus, an algorithm which does not rely on the camera calibration data was developed. This algorithm uses the similarity index of the two images to control the vergence angle. Due to the fact that the similarity index is a global characteristic, there is no need for accurate calibration of the cameras. Moreover, the detailed extraction of the disparity throughout the image is avoided, significantly decreasing the computational time and providing better results at the same time.

In this paper we attempt a comparative study of the most significant correlation measures for the direct

[1] Supported by the EU funded project Rescuer FP6–IST–2003–511492

ISBN 978-961-243-067-2

estimation of the vergence angle. In section 2, we appose the most significant measures such as the Sum of Absolute Differences (SAD), the Sum of Squared Differences (SSD), and the Normalized Cross Correlation (NCC), as well as the zero-mean counterparts of these measures. In Section 3, we compare time performance of these measures. In order to evaluate the performance of each measure, we extracted the disparity maps from our test images at the frames where the measures indicated verging and we examined whether in this situation the disparity value is close to zero. In the last part of Section 3 we prove that the performance of the best measure is independent of the image size. This is achieved by calculation of the MSE and the NMSE of the correlation indexes between the initial and sub-sampled images. Finally, in Section 4, we present the conclusions of our study.

2. SIMILARITY MEASURES

A summary of the difference and correlation measures is shown at Table 1. The simplest similarity measures are based on the SAD or the SSD, which is zero for identical ones. More computationally expensive measures, such as NCC, obtain values in the interval [-1, 1], with 1 standing for identical images, while ZNCC take values between [0, 2], where 0 stands for the highest correlation value.

Since we treat real images that have been captured under everyday surroundings, we noticed prominent illumination changes and significant contrast variations. The use of zero-mean versions of simple similarity measures (ZSAD, ZSSD, ZNCC) tend to smooth these discrepancies and, thus, they provide enhanced performance in our results. However, these measures increase the computational burden. Also it is apparent that the computational complexity increases according to the image size.

3. EXPERIMENTS

In this section, we compare the performance of the similarity measures for several real scenes. These are a pair of uncalibrated stereovision images. The right image of some representative pairs is presented in Figure 1. The test images have distinguished variations both in lighting and in content of the scenes. Image (*a*) contains different objects placed in such a way that covers the entire vergence space. Image (*b*) shows a common office environment with a person standing close to the cameras. Thereby we can examine the vergence behaviour of the robot head when an obstacle blocks a large region of its visual field. The last image represents a common office environment with no significant vergence targets along with the visual field of the robot head.

3.1 Comparison Methodology

In order to assess the optimum vergence angle we recorded the sequence of frames while verging from parallel position to the centre of the scenes mentioned in Figure 1. Each of these frames corresponds to a different vergence angle. The purpose of the experiments is to show that all similarity indexes have global extremum at the

Table 1. Difference and Correlation measures

Similarity Measures	Definition
Sum of absolute differences	$\sum_{(u,v)\in W} \left\| I_1(u,v) - I_2(u,v) \right\|$
Zero-mean sum of absolute differences	$\sum_{(u,v)\in W} \left\| \left(I_1(u,v) - \overline{I_1} \right) - \left(I_2(u,v) - \overline{I_2} \right) \right\|$
Sum of squared differences	$\sum_{(u,v)\in W} \left(I_1(u,v) - I_2(u,v) \right)^2$
Zero-mean sum of squared differences	$\sum_{(u,v)\in W} \left(\left(I_1(u,v) - \overline{I_1} \right) - \left(I_2(u,v) - \overline{I_2} \right) \right)^2$
Normalized cross correlation	$\dfrac{\sum_{(u,v)\in W} I_1(u,v) \cdot I_2(u,v)}{\sqrt{\sum_{(u,v)\in W} I_1^2(u,v) \cdot \sum_{(u,v)\in W} I_2^2(u,v)}}$
Zero-mean normalized cross correlation	$1 - \dfrac{\sum_{(u,v)\in W} \left(I_1(u,v) - \overline{I_1} \right) \cdot \left(I_2(u,v) - \overline{I_2} \right)}{\sqrt{\sum_{(u,v)\in W} \left(I_1(u,v) - \overline{I_1} \right)^2 \cdot \sum_{(u,v)\in W} \left(I_2(u,v) - \overline{I_2} \right)^2}}$

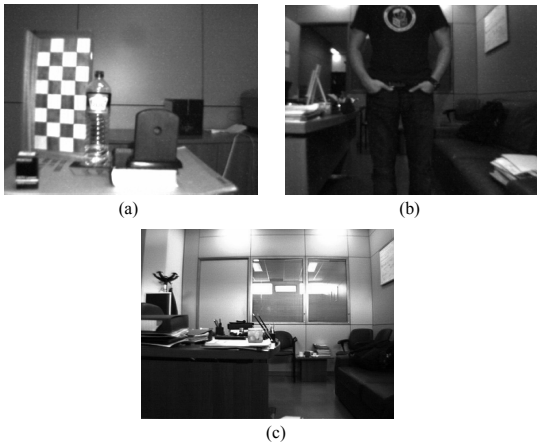

Fig. 1. Test images (only the right one).

desirable vergence angle, and, thus, a comparison of these according to their time performance is straightforward. We have examined the performance of these measures for various image sizes ranging from 640x480 to 20x15. Figure 2 shows the plots of all indexes, against the vergence angle, for the test image (*a*) and for all their respective sub-sampled ones.

The experiment shows that the global extremum appears at the expected vergence angle while it remains unaffected, as we reduce the dimensions of the test images. In spite of the fact that the difference measures (SAD, SSD) and the zero mean versions of them (ZSAD, ZSSD) find the right frame and vergence angle, they have shown poor performance in all test images. More specifically, the global minimum is not clearly identified on the plot, especially on the sub-sampled versions. On the other hand, the correlation measures give a clear maximum or minimum, for the NCC and ZNCC respectively, at the frame where the correct vergence angle occurs. Additionally, we can obviously notice from the plots *e* and *f* in Figure 2 that the plot of ZNCC has a smoother outline than the plot of NCC, without exhibiting major variations between the original and the reduced images. The combination of this robust performance and its stability to environmental changes render the ZNCC to outperform the other compared measures.

The same performance of the measures was observed in all the other tests we carried out. In these tests the measures show that the best match occurs at the same frame where we observe the best vergence angle. We repeated the experiments reducing gradually the image size down to 40x30 obtaining the same results. For smaller images sizes though, the majority of the measures, including the ZNCC, failed to indicate the correct vergence angle. Figure 3 demonstrates the respective plots of the measure ZNCC for the test images *b* and *c*. We can notice that the vergence angle is estimated accurately, even in cases where the background is cluttered and, thus, the ambiguity is high (Figure 4(b)).

During the above described experimental procedure, the computational time was recorded and the results are presented in Figure 4. As it is clearly illustrated the computational time decreases exponentially according to the image size, no matter to the measure used. As we expected, the measure SAD, which has the lowest computational complexity, needs the least time to find the correct vergence angle, whilst the ZNCC, which is the most complex measure, is the most time consuming. Especially for the initial dimensions of the test images the difference between their time performances is 20 ms. However, the lower the resolution of the images, the smaller the time difference. Thus, for 40x30-pixel image size it is tumbled to 0.1 ms. This negligible difference in time performance in combination with the great smoothness in the plots presented in Figures 2f and 3, cause the ZNCC measure to be our suggestion, in order to achieve accurate and reliable results for real-time vergence angle estimation using image sizes of 40x30 pixels.

3.2 Evaluation Methodology

In order to study how the chosen image size (40x30) might affect the accuracy of the vergence estimation method, for the suggested (ZNCC) index, we calculated the MSE and NMSE between the original image (640 x480) and the chosen one. The estimation of the mean errors was done using the formulas below:

$$MSE = \frac{1}{n}\sum_{i=0}^{n}\left(f_{640}(i) - f_a(i)\right)^2 \tag{1}$$

$$NMSE = \frac{\sum_{i=0}^{n}\left(f_{640}(i) - f_a(i)\right)^2}{\sum_{i=0}^{n}\left(f_{640}(i)\right)^2} \tag{2}$$

where f_a is the curve of a sub-sampled image and n stands for the number of the recorded frames.

We have estimated the MSE and the NMSE between the original image and the all the sub-sampled down to 20x15 pixels. The results are exhibited in Table 2, where we can notice that the error (either calculated as MSE or as NMSE) never exceeds 1‰. This is another indication that we can use sub-sampled images for the vergence control, without a significant loss of information.

In active stereo vision systems, when we verge appropriately, the disparity between the left and the right image approaches very low values. Therefore the best vergence angle appears when the disparity of

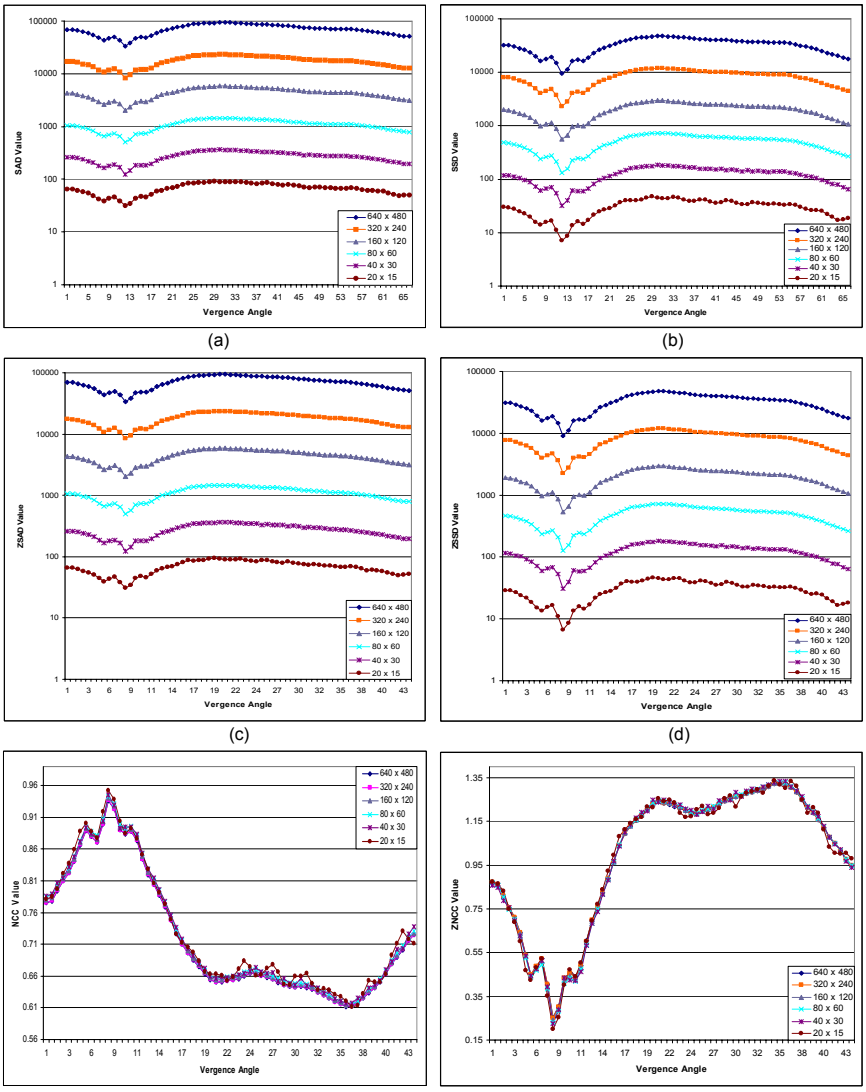

Fig. 2. Similarity indexes against the vergence angle for test image a. The lines correspond to the different dimensions. The minimum values represent the correct vergence angle except the NCC index where the maximum value corresponds to the highest correlation.

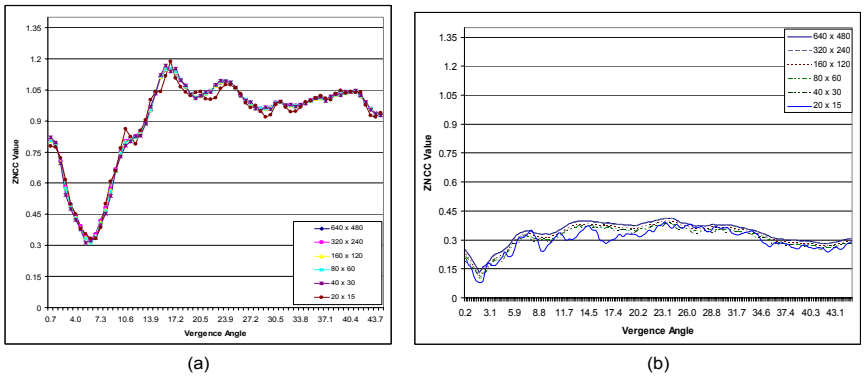

Fig. 3. ZNCC indexes of the test image b (left) and test image c (right)

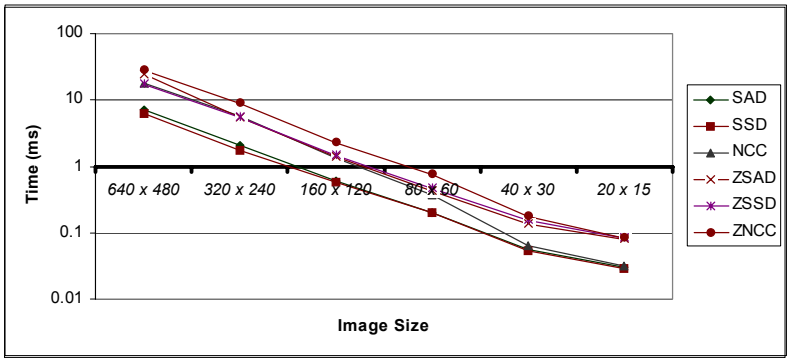

Fig. 4. Time performance for different image sizes and for each measure

the images is very close to zero. Thus, in order to evaluate the choice of the ZNCC measure for the estimation of the vergence angle, we extracted the respective disparity maps. These were computed using (Kotoulas *et al.,* 2005). This method employs SAD and, in order to avoid the production of a noisy output, the images are pre-processed using a one dimension weighted mean filter and post-processed by a cellular automata filter. In Figure 5(a) and (b) we illustrate the resulting disparity maps for the case of a common office environment with cluttered background, when the cameras are in parallel position and when they converge, respectively. It is apparent from the respective histograms that when the optimum vergence angle occurs, the value of disparity is zeroed.

Table 2. MSE and NMSE of the ZNCC index for all test images.

Test image *a*	MSE (%)	NMSE (%)
$f_{640} - f_{320}$	0.000039	0.000035
$f_{640} - f_{160}$	0.003198	0.002937
$f_{640} - f_{80}$	0.012091	0.011105
$f_{640} - f_{40}$	0.017241	0.015834
$f_{640} - f_{20}$	0.063771	0.058568

Test image *b*	MSE (%)	NMSE (%)
$f_{640} - f_{320}$	0.000022	0.000026
$f_{640} - f_{160}$	0.001398	0.001630
$f_{640} - f_{80}$	0.004120	0.004803
$f_{640} - f_{40}$	0.017162	0.020007
$f_{640} - f_{20}$	0.060502	0.070531

Test image *c*	MSE (%)	NMSE (%)
$f_{640} - f_{320}$	0.001280	0.010965
$f_{640} - f_{160}$	0.027581	0.236270
$f_{640} - f_{80}$	0.071940	0.616261
$f_{640} - f_{40}$	0.099250	0.850206
$f_{640} - f_{20}$	0.209395	1.793728

4. CONCLUSIONS

In this paper, we compared the performance of several similarity measures and their use in calculating the vergence angle of an active vision stereo head. The measure that exhibits the smoother performance, even in the most difficult scenes, was the ZNCC. This measure was proved to be fast, accurate and robust to environmental changes, as it compasses the correct vergence angle even with

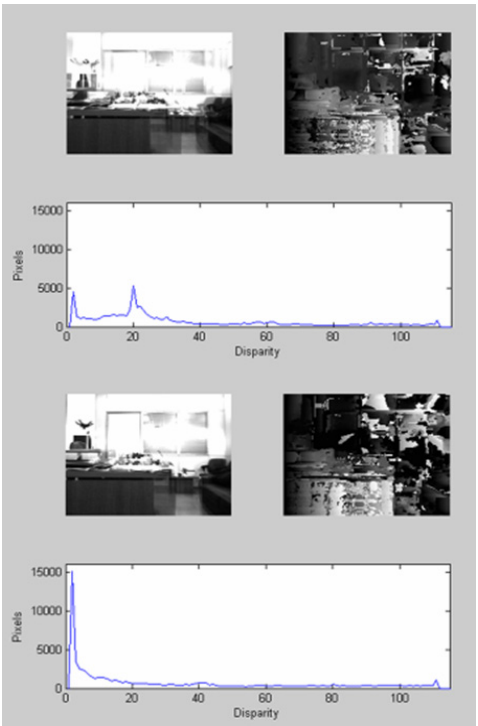

Fig. 5. Evaluation results for the case presented in Figure 2(c). On the upper side the fused images, the extracted disparity maps and the histograms of the disparity map are shown when the cameras are fully paralleled. On the lower side the same are shown for fully verged cameras. We can clearly notice that when the best vergence angle occurs, the majority of the disparities are close to zero.

extremely low-resolution images. Furthermore, we showed that for a stereo scene we can attain the best vergence angle without using sophisticated methods such as log–polar mapping because we are able to use successfully the ZNCC with cartesian images of 40x30-pixel size. In order to verify the precision of our recommendation we extracted the respective disparity maps. We showed that when the best vergence angle occurs the value of the majority of the disparities is zeroed. Also, in order to examine the similarity between the resulting ZNCC indexes and the original, we calculated the MSE and the NMSE. The error never exceeds 1‰, even in difficult and cluttered scenes. Thus, the proposed implementation can underlie an excellent approach for the construction of a reliable and accurate real time stereoscopic robot vision system.

REFERENCES

Arz G., András T., Daniel Bratanov, Nikolay Zlatov, Roberto Guzmán Diana, Stefko Burdzhiev, Antonios Gasteratos, David Pereira, Carlos Beltrán, and Uberto Delprato. (2005). "EU FP6 Project RESCUER Develops Intelligent Robot and Information Technologies for EOD/IEDD/Rescue Mission." *1st Midlle-East European Conference on Military Technology*, Budapest, Hungary

Bernardino, A., and Santos-Victor, J. "Correlation based vergence control using log-polar images." *In Proc. of the 4th Int. Symposium on Intelligent Robotic Systems*, Lisbon, Portugal.

Diaz, J., Ros, E., Carrillo, R., and Prieto, A. (2007). "Real-Time System for High-Image Resolution Disparity Estimation." *Image Processing, IEEE Transactions on*, 16(1), 280-285.

Forstmann, S., Kanou, Y., Jun, O., Thuering, S., and Schmitt, A. "Real-Time Stereo by using Dynamic Programming." 29-29.

Hansen, M., and Sommer, G. (1997). "Real-time Vergence Control using Local Phase Differences." Computer Science Institute, Christian-Albrechts-University Kiel.

Hee-Jeong, K., Myung-Hyun, Y., and Seong-Whan, L. "A control model for vergence movement on a stereo robotic head using disparity flux." 491-494 vol.4.

Kotoulas, L., Gasteratos, A., Sirakoulis, G. C., Georgoulas, C., and Andreadis, I. "Enhancement of Fast Acquired Disparity Maps using a 1-D Cellular Automation Filter." *Proceedings of the Fifth IASTED International Conference on Visualization, Imaging, and Image Processing*, Benidorm, Spain.

Manzotti, R., Gasteratos, A., Metta, G., and Sandini, G. (2001). "Disparity Estimation on Log-Polar Images and Vergence Control." *Computer Vision and Image Understanding: CVIU*, 83(2), 97-117.

Niitsuma, H., and Maruyama, T. (2004). *Real - Time Detection of Moving Objects*, Springer Berlin / Heidelberg.

Sang-Bok, C., Bum-Soo, J., Sang-Woo, B., and Minho, L. "Trainable attention model based vergence control for active stereo vision system." *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*, 519-524.

Sun, C. (2003). "Fast algorithms for stereo matching and motion estimation." ustralia-Japan Advanced Workshop on Computer Vision, Adelaide, Australia.

Tsang, E. K. C., and Shi, B. E. "Active Binocular Gaze Control Inspired by Superior Colliculus." *International Joint Conference on Neural Networks. IJCNN '06.* , 7-14.