MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

Disparity Selection in Binocular Pursuit

Atsuto Maki, Tomas Uhlin and Jan-Olof Eklundh

Computational Vision and Active Perception Laboratory (CVAP)
Royal Institute of Technology (KTH), S-100 44, Stockholm, Sweden

Abstract: *This paper presents a technique for disparity selection in the context of camera vergence in binocular pursuit. For vergence control in binocular pursuit, it is a crucial problem to find a disparity which is corresponding to the target among multiple disparities generally observed in a scene. To solve the problem of the selection, we propose an approach based on histogramming the disparities obtained from a phase-based disparity estimation algorithm. The idea is to slice the scene using the disparity histogram so that only the target remains. The slice is chosen around a peak in the histogram, a peak which is found by keeping track of the peak as it moves in time, using prediction of the targets disparity and location in the image. The tracking of the peak enables robustness against any dominant objects in the scene. The approach is investigated through experiments and shown to work appropriately.*

Keywords: disparity, vergence, phase-based algorithm, disparity selection, binocular pursuit

1 Introduction

Pursuit of a target using a stereo camera system requires mechanisms for version to track the target on the horopter, and vergence to fixate it in depth (see Fig.1). Since those two mechanisms are orthogonal to each other, pursuit can be performed by controlling them independently. The research described here is concerned with the mechanism for vergence. To keep the cameras verged on a target, the vergence system must measure the current vergence error. We utilize binocular disparity as the most useful visual cue to vergence, because of the straightforward mapping into vergence error. A crucial problem in applying this together with pursuit is to estimate the disparity corresponding to the target, while multiple disparities are observed due to other objects in the scene. I.e., a selection process has to take place among the multiple disparities observed in the image.

The objective of the work presented here is to solve the problem of disparity selection. Typical for this problem is that not only do we want to estimate a disparity at each instance of time, but this disparity should also be consistent over time and with the target location, hence disparity selection is tightly coupled with pursuit. As the disparity estimator a method which is based on the output phase of bandpass filters [San88] [WK89] is employed. The advantage of the method is in the computational cost, the stability against varying lighting condition and especially in the direct localization of the disparity estimation. Provided with a disparity map from the phase-based algorithm, we introduce an approach to select the disparity of the target by forming a histogram and selecting the corresponding peak. The performance of the whole scheme is investigated through experiments.

The article comprises of Section 2, briefly describing the phase-based algorithm, and Section 3, where the method of disparity selection is presented, followed by the experiments shown in Section 4. A conclusion is given in Section 5. The presented techniques are part of the framework (see Fig.2) which is to perform real-time smooth pursuit on our vision system, the KTH-head [PE92].

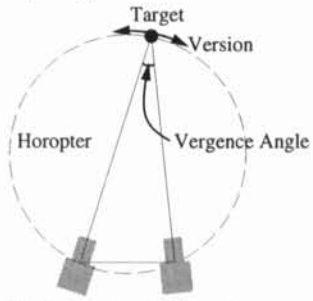


Figure 1: Geometry of binocular pursuit.

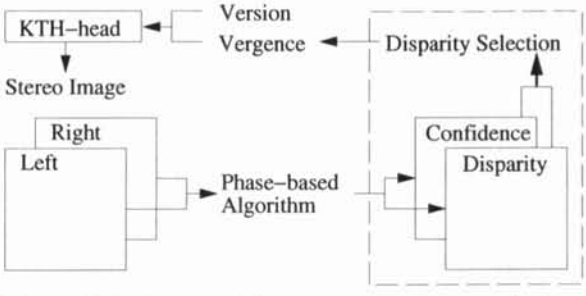


Figure 2: A diagram of the framework. Presented in the paper is the part marked by the dashed line.

2 Phase-based disparity estimation

The fundamental concept of the phase-based approach is to convolve the left and right stereo images with a complex filter, such as a Gabor filter, and then estimate the local disparity by computing the complex phase difference of the filter output. As the local shift between stereo images is linearly proportional to the local phase difference, a disparity estimate is obtained at each point in the image.

Disparity map
Let a complex filter be convolved with the left and right images respectively and produce $V_l(x, y)$ and $V_r(x, y)$. Since the computation of the convolution in Fourier domain has contribution mainly from the neighborhood of the central filter frequency ω[radian/pixel], the convolution products are approximately related to each other by phase shift, which arises from the spatial shift (i.e. disparity):

$$V_l(x, y) \approx e^{j\omega D(x,y)} \cdot V_r(x, y). \quad (1)$$

$D(x, y)$ denotes disparity at (x, y) in the image. The relation then leads to the disparity through the computation of the complex phase difference:

$$D(x, y) \approx (arg[V_l] - arg[V_r])/\omega. \quad (2)$$

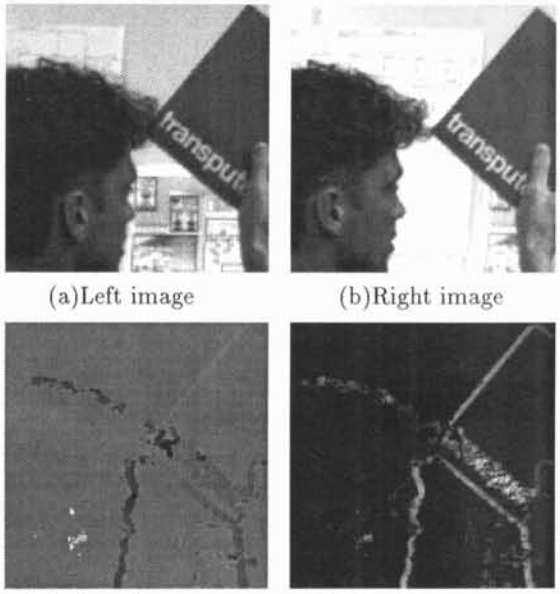
This is strictly valid only for filters of infinitesimal bandwidth, arising directly from the Fourier shift theorem[San88].

Confidence map
In order to check the feasibility of the estimated disparity and threshold unreliable estimation, we also compute a confidence value defined on the basis of the magnitude of the convolution product:

$$C(x, y) \equiv \frac{mag[V_l] \times mag[V_r]}{mag[V_l] + mag[V_r]}. \quad (3)$$

Complex filter
While the complex filter has to satisfy several constraints and several different filters have been proposed accordingly [Wes92], we employ discrete approximations to the first and second derivatives because of their computational simplicity, that is $(-1, 0, 1)$ and $(1, 0, -2, 0, 1)$, thus sacrificing accuracy. The frequency of the filter is regarded as $\omega = \pi/2$[radian/pixel]. The maximum disparity which a filter can accurately determine is limited to one-half the wavelength of filter. To handle larger disparities, a coarse-to-fine technique can be employed [MUE93].

Fig.3 shows an example of a disparity and confidence map that is used in the disparity selection in the next section.



(a)Left image (b)Right image

(c)Disparity map (d)Confidence map

Figure 3: Example of a disparity and confidence map. (a)(b) A sample stereo image pair. The fixation point is in the center around the corner of the book. (c) The disparity map. Estimation is valid only where the corresponding confidence is above a threshold. The higher the gray scale is, the larger the disparity estimation is (further away in the scene). (d) The confidence map. The higher the gray scale is, the higher the confidence value is.

3 Disparity selection

In a scene where several objects exist at different depths, multiple disparities are observed. The objective here is to select the one that is corresponding to the target in binocular pursuit. Other reserchers have reported related work, e.g. zero-disparity filtering together with cepstral-filtering and vergence control in smooth pursuit [CB92], also disparity estimation by cepstral-filtering and back projection for disparity selection in vergence on objects in static scenes [TOM94]. For the specific purpose of binocular pursuit, it is desirable to have an approach which utilizes both binocular and monocular ques in time, and is stable regardless of the size and absolute location of the target. We also argue that we can sacrify accuracy in the disparity estimation and in turn gain speed, in which accurate vergence and pursuit is obtained quickly in time rather than in one shot. The selection process presented below, together with a parallell motion estimation process, seems to be promising in these respects.

Histogram-based selection
Based on a disparity map, $D(x, y)$, and a confidence map, $C(x, y)$, we compute a histogram $H(D_d)$ with

respect to the discrete disparities D_d:

$$H(D_d) = \sum_{x,y} C(x,y) \text{ for } \{(x,y) \mid D_d <= D(x,y) < D_{d+1}\}. \quad (4)$$

$H(D_d)$ is defined as the sum of the confidence values at the pixels where the disparity is estimated to D_d. Multiple peaks therefore appear in the histogram corresponding to the different disparities[1]. Given a prediction of what disparity the target should have, the closest peak can be selected as the estimate of the target disparity.

Fig.4 (a) is an example of such a histogram, produced based on the disparity and confidence map shown previously in Fig.3. Four large peaks are observed in the histogram corresponding to different parts of the image lying at different depths; the head of the man, the hand, the lower and the upper part of the book. A neighborhood around a peak[2] in the disparity histogram is back projected into the sample image, see Fig.3 (b). The selected peak turned out to arise from the lower part of the book.

Prediction

An important feature of the disparity selection process is the disparity prediction, that is, given an initial disparity of the target the consequent selection at each instant in time is based on a prediction of the disparity. For computational simplicity a linear predictor is used with a weighting factor α[3]:

$$D_{p(k+1)} = D_{s(k)} + P_{(k)} \quad (5)$$

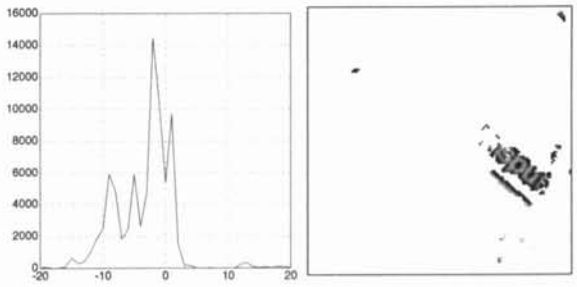
$$P_{(k)} = \alpha \cdot (D_{s(k)} - D_{s(k-1)}) + (1-\alpha) \cdot P_{(k-1)} \quad (6)$$

$D_{s(k)}$ and $D_{p(k)}$ represent the selected and predicted disparity at frame number k while $P_{(k)}$ denotes the predicted change. A crucial problem is the case when the peak of interest disappears temporarily in time. This could happen in a scene where more than one object simultaneously come close in depth. To handle such a situation, information about the location of the target is incorporated in the scheme. One could try to use location information provided by the selection process itself, but the approach we are pursuing utilizes a parallell motion estimation process. However, this work is outside the scope of this paper. We only mention briefly that the previously selected target location provide areas for motion estimation, and that the subsequent selection of disparity should give consistent estimations of target location.

[1] A threshold is used to eliminate small peaks due to noise.

[2] The size of the slot taken around the selected peak is 2 pixels, which corresponds to several centimeter in the distance.

[3] $0 \leq \alpha \leq 1$, α is set to be 0.2 in the experiments.



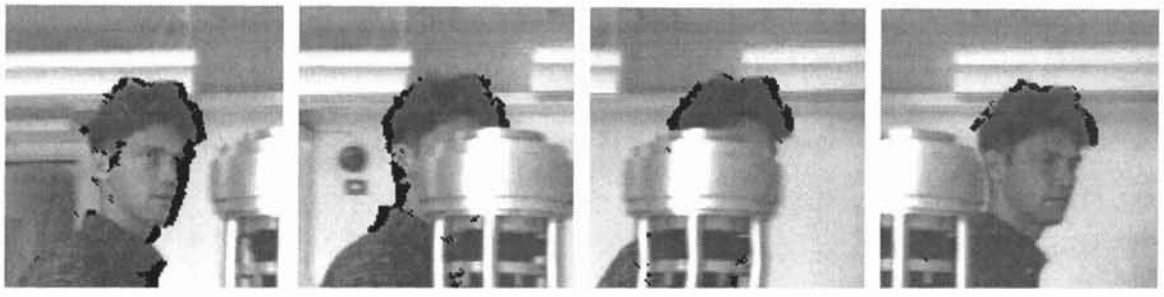
(a)Histogram (b)Back projection

Figure 4: Example of disparity selection. (a) Histogram. Horizontal axis for disparity estimates. Vertical axis for sum of the confidence values. (b) Back projection of lower part of the book.

4 Experiments

Experiments have been performed on some image sequences so as to demonstrate the behavior of the presented scheme over time. Two differnt sequences have been taken by a binocular camera system whose gaze direction was controlled manually while the vergence angle was fixed. Fig.5 shows a sequence where the camera tracks a person passing behind an object. The head of the man is the target, which is kept to be in the center throughout the sequence. Fig.6, on the other hand, shows the situation where the camera gets stuck on the object occluding the target. This is to show that if a system for some reason would get stuck on a occluding object, the disparity selection process would have the ability to cope with such a situation, and thus a complete system including disparity selection could deal with such a situation. The frame rate is 40 msec in both cases one image is shown for every tenth frame.

With the initial disparity corresponding to the target, in both experiments, the histogram is generated at each frame in the sequences and disparity selection takes place accordingly. Marked dark is the part where the disparity estimation is around the selected peak. The edge on the face in Fig.5 (d) Fig.6 (a) is not marked since that part is occluded in the other image. Dark marks are also observed in irrelevant parts of the image due to erroneous disparity estimation. As a whole, however, the selected disparity remains on the target over time. Note especially how the selected disparity still sticks to the target even when the target is mainly occluded, or has halfaway disappeard out of the image (see Fig.6 (c)(d)). A simple selection of the highest peak would not do (see Fig. 7), and surely many, if not all, of todays monocular pursuit systems would fail.

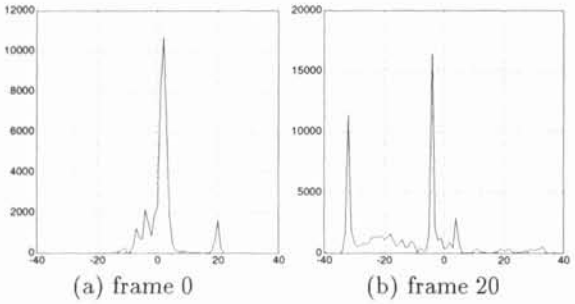


(a) frame 0 (b) frame 10 (c) frame 20 (d) frame 30

Figure 5: Experiment I. The target is kept to be in the center. Back projection on the right input image is marked dark. 1 frame = 40 msec.

(a) frame 0 (b) frame 10 (c) frame 20 (d) frame 30

Figure 6: Experiment II. The target in the center is taken over by a occluding object. Back projection on the left input image is marked dark. 1 frame = 40 msec.



(a) frame 0 (b) frame 20

Figure 7: Histograms in Experiment II. (a) The target corresponds to the large peak. (b) The target corresponds to the small peak to the right of the large.

5 Conclusion

In this article, we have considered the problem of disparity selection, a vital component of successful camera vergence in binocular pursuit, in which also the incorporation of time in disparity estimation plays a central role in the success. Based on the disparity and confidence maps produced by the phase-based algorithm, a histogram-based approach utilizing a peak predictor has been introduced. Through the experiments it can be seen that the scheme works well, especially in terms of the robustness during occlusion. Future work will be directed (i) to improve the performance of the disparity estimator, (ii) to cope with multiple objects with the same disparity, and (iii) to implement the scheme in real-time.

References

[CB92] D. Coombs and C. Brown. Real-time smooth pursuit tracking for a moving binocular robot. In *Proc. CVPR*, pages 23–28, 1992.

[MUE93] A. Maki, T. Uhlin, and J. O. Eklundh. Phase-based disparity estimation in binocular tracking. In *Proc. 8th SCIA*, pages 1145–1152, May 1993.

[PE92] K. Pahlavan and J. O. Eklundh. A head-eye system - analysis and design. *CVGIP:Image Understanding*, 56(1):41–56, July 1992.

[San88] T. D. Sanger. Stereo disparity computation using gabor filters. *Biol. Cybern.*, 59:405–418, 1988.

[TOM94] J. Taylor, T. Olson, and W. N. Martin. Accurate vergence control in complex scenes. In *Proc. CVPR*, pages 540–545, 1994.

[Wes92] C.-J. Westelius. Preattentive gaze control for robot vision. *Tech. Lic. Dissertation, Computer Vision Lab., Linköping University, Sweden*, (LiTH-ISY-I-1327), June 1992.

[WK89] R. Wilson and H. Knutsson. A multiresolution stereopsis algorithm based on the Gabor representation. *Proc. IEE International Conf. Im. Proc. and Applic., Warwick. U.K.*, pages 19–22, 1989.