논문 98-7-6-03

A High Speed Vision Algorithms for Axial Motion Sensor

Stephane. MOUSSET*, Pierre. MICHE, Abdelaziz. BENSRHAIR*,. Sang-Goog LEE**,

ABSTRACT

In this paper, we present a robust and fast method that enables real-time computing of axial motion component of different points of a scene from a stereo images sequence. The aim of our method is to establish axial motion maps by computing a range of disparity maps. We propose a solution in two steps. In the first step we estimate motion with a low level computing for an image point by a detection estimation-structure. In the second step, we use the neighbourhood information of the image point with morphology operation. The motion maps are established with a constant computation time without spatio-temporal matching.

I. Introduction

The determination of 3-D motion is an area of very active researches. This is explained by the fact that it is an important task in mobile robot control and car driving assistance.

Most of approaches consist in two main steps: at first, objects of the scene are isolated ; then, their motion components are estimated [1][2][3].

Sabata and Aggarwal [4] show that motion transformation has a well defined structure in terms of various components. When computing a motion transformation, this structure has to be taken into account and all physical constraints have to be satisfied. In general case, these constraints are non-linear. Hence, motion estimation usually turns out to be a non-linear optimisation problem. The solution to such constrained non-linear optimisation problem is not simple. Sabata and Aggarwal propose a decomposition of motion parameter computation in four steps:

1 - selsecting the feature to be used for the task,
2 - representing the motion transformation
3 - representing the feature,
4 - using the representation to compute the transformation.

Several researchers has investigated the possibility of recovering the 3-D motion without point to point correspondence. Lin et al[5] present some eigenstructure-based algorithms for estimating the motion parameters of a rigid object from scaled orthographic projections. Li and Duncan [6] present a method to determine 3-D motion and 3-D structure in case of a motion restricted to a translation. They suggested a two-step procedure applied to a sequence of stereoscopic images. In the first step, translational motion parameters are determined in an independent way in the two images. In the second step, the binocular flow information is used to find matching correspondence.

* PSI-LCIA, INSA de Rouen FRANCE
** Sensor Technology Resear Center, Kyungpook National University
<접수일자 : 1998년 4월 30일>

The specificity of our approach lies in two points: first, objects are not isolated before motion analysis, and then, we only consider the axial motion transformation.

In a first section, we present the special stereo image system. Then we study the determination of the axial motion component of points. In a third section, we explain the method to compute the axial motion map from a disparity map sequence. Then, we present some problems concerning spatio-temporal object tracking and noise. In the last section we present some experimental results and conclusion.

II. Configuration of stereo image system

To obtain the three-dimensional information, at first we need a dual camera model. In most cases, the modelling of two camera implies to solve a set of non-linear equations involving a lot of coefficients. So, in order to simplify the problem of camera calibration as well as feature extraction and feature matching, we chose a spatial stereo system which allows the stereo process to be quicker and the algorithms to be reduced to a one-dimensional problem.

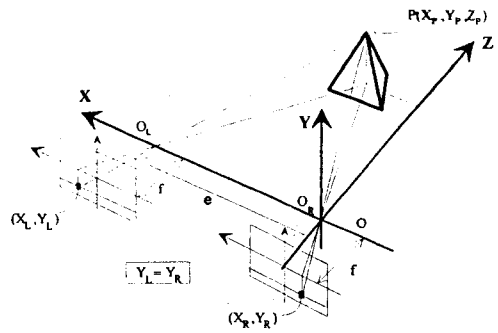

Figure 1. Spatial camera configuration

In this configuration (figure 1), we suppose that the two cameras have parallel optical axes with equal focal lengths and the baseline is horizontal and parallel to the scan lines of images. Furthermore the Z-axis is coplanar with the optical axes of the cameras, so all epipolar lines are parallel to the baseline and they are confused with horizontal scan lines of images.

Given this camera model, we search the possible matching feature points in only one dimension, i. e. only in two image lines (right and left) having the same Y-axis value, rather than in the whole image. The calculation formulas of an object point $P(X_P, Y_P, Z_P)$ are :

$$X_p = -\frac{e}{\delta_z} \cdot X_R$$
$$Y_p = -\frac{e}{\delta_z} \cdot Y_R \qquad (1)$$
$$Z_p = \frac{e \cdot f}{\delta_z}$$

with X_R, Y_R, X_L, Y_L the local coordinates respectively in the left and in the right images, e the length of the baseline, f the focal length of the two cameras, and d_z the diparity equal to $\delta_z = X_L - X_R$.

Using this configuration, we have developed original stereovision algorithms [7] [8] to compute in a short time disparity maps . The motion analysis that we present below, is based on a sequence of such disparity maps.

III. Axial motion determination

The aim of this section is to estimate the axial motion component of a tracked point from a long sequence of stereo images.

III-1. Axial motion relation

Consider a point at distances $ZP\ (t)$ and $ZP\ (t+t)$ respectively at times t and $t+t$. The axial motion component is given by a Taylor's development to the first order of the depth value.

$$Z_P(t+\Delta t) = Z_P(t) + \dot{Z}_P(t) \cdot \Delta t + o(\Delta t) \qquad (2)$$

where $\delta t = (t_2 - t_1) \rightarrow 0$

Assuming the axial velocity to be constant during Dt and using the relation (1), then, it can be expressed by:

$$\dot{Z}_p(t) = \frac{K}{\Delta t}\left(\frac{1}{\delta_z(t+\Delta t)} - \frac{1}{\delta_z(t)}\right) \quad (3)$$

where K is a parameter which depends on the characteristics of the image acquisition system.

For a long sequence, assuming the motion is constant, the estimation of axial motion component $W(t,n)$ is determined by the following recurring relation:

$$W(t+\Delta t, n+1) = W(t, n) + K_G(\dot{Z}_p(t+\Delta t) - W(t, n)) \quad (4)$$

where n is the length of the sequence and K_G is the ratio of correction due to the observation at the instant $(t+\Delta t)$. For our configuration we have $K_G = 0.5$. For the first measure we using this initial condition:

$$W(t, 0) = \dot{Z}_p(t) \quad (5)$$

With the relations (4) and (5), the axial motion component may be write in a other form:

$$W(t, n) = \frac{1}{2^{n+1}} \cdot \dot{Z}_P(t - n \cdot \Delta t) + \sum_{i=0}^{n} 2^{n+1} \cdot \dot{Z}_P(t - i \cdot \Delta t) \quad (6)$$

This compute of estimation of axial motion component is very fast and the estimation converge rapidly to the correct value.

III-2. Convergence of relation

Consider the difference $W(t,n) - W(t,n-1)$. When n tends to infinity, this difference is maximised by $\sigma(t, n, \varepsilon)$ as defined by equation (7):

$$\sigma(t, n, \varepsilon) = \frac{K}{\Delta t}\left(\frac{1}{2^n(\delta(t - n \cdot \Delta t) - \varepsilon)}\right) \quad (7)$$

where is the error on disparity. The convergence of relation (6) is established by the fact that $\sigma(t, n, \varepsilon)$ approaches zero when n tends to infinity.

III-2. Error analysis

Errors in axial motion component $W(t,n)$ are due to the following causes:

- inaccurate calibration,
- disparity error.

On one hand, inaccurate calibration involves many complicated factors whose effects are very difficult to analyse rigorously. However, experiments with real image sequences (presented below), show that our method limits miscalibration effects. On the other hand, disparity error causes an error in axial motion $\Delta\dot{Z}_P$ which can be modelised by an additional white Gaussian noise with a null mean. For a constant axial motion component, the following relation is verified:

$$\lim_{n \to \infty} \sum_{i=0}^{n} \Delta\dot{Z}_P(t - i \cdot \Delta t) = 0 \quad (8)$$

The uncertainty of recurrent relation (6) can be maximised by the one of $\dot{Z}_P$ for the last acquisition time:

$$\frac{\Delta\dot{Z}_P}{\dot{Z}_P} = \frac{(\delta(t) + \delta(t+\Delta t)) \cdot \Delta\delta}{\delta(t) \cdot \delta(t+\Delta t)} \quad (9)$$

where $\Delta\delta$ is set to 1. The uncertainty of axial motion decreases as disparity increases.

IV. Axial motion map computation

VI-1. Implementation of the motion algorithm

The aim of our method is to establish axial motion maps with a range of disparity maps. Disparity maps are obtained with fast and self-adaptive algorithms from a stereo image pairs. We assumed that the disparity maps are determined in real-time with a frequency of 5 images by second. These images are computed on a pipeline

structure as described by the figure presented below.

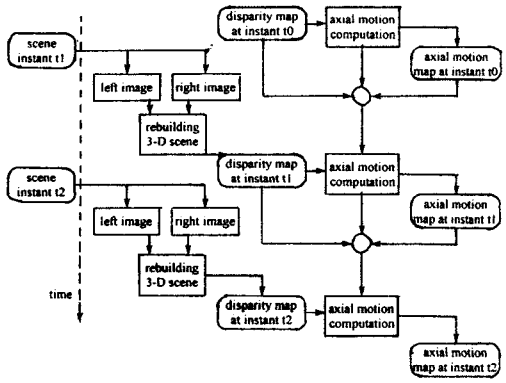

Figure 2. Structure of the algorithm

In this figure, the "rebuilding 3-D scene" bloc is the compute of the distance of characteristic point in two stage [9]. The first stage is a contour extraction for each image, the second stage is the stereo matching.

The "axial motion computation" bloc consists of two steps. In the first step, we estimate motion with a low-level computing for an image-point by a detection-estimation structure presented in figure 3.The aim of this procedure is to compute axial motion without region matching. In the second step, we use the neighbourhood information of the image-point for the morphology operation described in spatio-temporal tracking chapter. The neighbourhood information is about grey-level, axial motion component, state of tracking of the neighbourhood points.

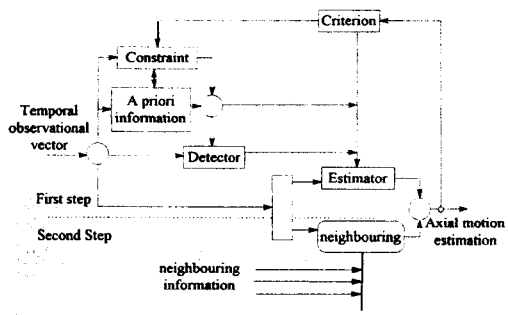

Figure 3. Axial motion computation structure

VI-2. Detection-Estimation Structure

This structure is constructed on the principle of detection-estimation with defined criterion and constraints. The estimation module is defined with the constraint that axial motion of objects are constant as described by (6). The *a priori* information module define a maximum value of axial acceleration for all image-points. The criterion defines the maximum value of axial motion. It is computed for the whole motion map.

VI-3. Detection module

The detection module is the pivot of this structure. Its problematics can be summed up by:

1 - correct tracking of image point,
2 - appearance of object,
3 - disappearance of object,
4 - corrupted data presence.

These different issues are processed hierarchically in an order implied by security reasons. Thus, appearance of an object is processed in priority. We assumed that the context of the observed scene and evolution of object motion are known. Then the tracking evolution possibilities are limited. We can describe this tracking evolution by an automaton as following:

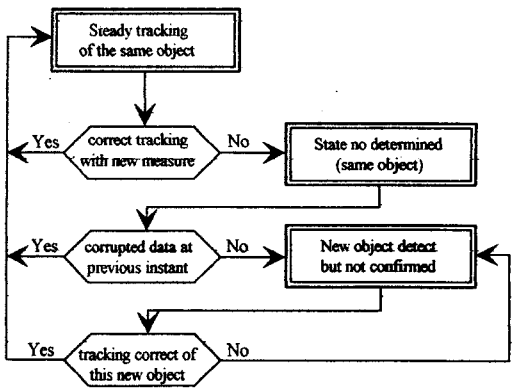

Figure 4. tracking evolution

There are tree possible states : correct tracking, state no determined and new object. For a pixel, the appearance or the disappearance of a object is a

identical problem. For the state no determined, we assume in first time the data is bad. So the compute of axial motion component uses previous data.

This procedure of tracking operates at the point-image level. It is fast and independent on the number of objects in the scene. Otherwise, the detector initialises the estimator parameters with regard to the detected case. Thus, when a new object is detected relation (6) is re-initialised.

VI-4. Spatio-temporal tracking.

In the second step, we do not want to use spatio-temporal matching in our sequences to respect our real-time constraint. The frequency of image acquisition being high, this allows to consider the hypothesis that object have little displacements between two acquisitions. Then, object images in two consecutive depth maps overlaps. The axial motion component of an object is evaluated in this overlapping area.

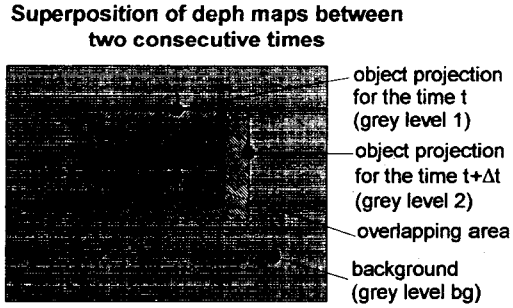

Figure 5. Object overlapping in depth maps

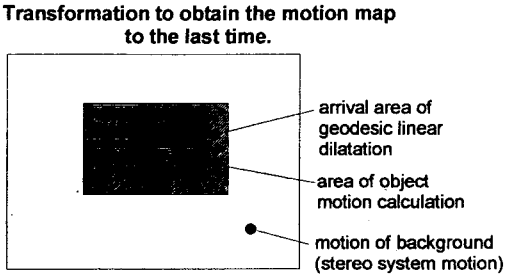

Figure 6. Geodesic linear dilation

This result has then to be extended to the totality of the projected surface of the last time. This operation is realised using a morphology operation which is a geodesic linear dilation. This operation is described in figure 5 and 6.

This procedure does not take into account the disappearance or appearance of an object. These problems are studied by the detection-estimation structure.

VI-5. Computation of motion for far objects.

In the case of far objects, an axial displacement does not necessarily implies variation of the disparity between two consecutive acquisitions. To solve this problem, the time interval used in the calculation of motion is not constant but is conditioned by a modification of the disparity.

V. Experimental results

The efficiency of our method is demonstrated by the following experiments with outdoor scene. This scene involves a car which is progressing towards the viewer. In photo 1 and 2, we present an example of stereo pairs obtained on successive timse.

In photo 3, the corresponding disparity maps are presented. These maps are presented using a grey level coding: points of high grey levels correspond to short distances. In photo 4, the computed axial

Photo 1. Stereo images(left and right) at time t=5.6s

Photo 2. Stereo images(left and right) at time t=5.8s

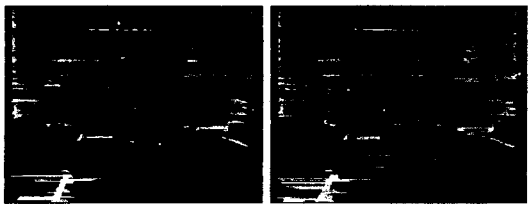

Photo 3. Disparity maps at time t=5.6s and t=5.8s

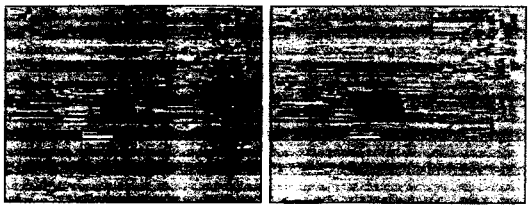

Photo 4. Axial motion component maps at time t=5.6s and t=5.8s

In photo 4, the computed axial motion component maps are shown. These maps present a different coding: motionless points have a medium grey level, points moving toward the viewer have higher grey levels and those moving away have lower grey levels.

At present, 128X128 pixels stereo images are processed is less than 0.5 s on an IPX SUN workstation.

V-1. Noise analysis

The stereo images are noisy. This noise which may be considered as gaussian and additive, affects disparity maps. This implies two types of defaults.

The first default is the non detection of characteristic elements (feature points). The computation of 3-D representation for these points is erroneous. A low-pass filter or a median filter improves disparity maps as it is the case for the disparity maps above presented. This corrupted data problem is processed by the detection-estimation structure.

The second default introduces an error of one pixel in feature points location which generates an inaccurate depth information. Currently, this problem is overcome efficiently by calculating the speed on a long stereo image sequence and by using the maximum axial motion constraint.

V-2. Speed evolution analysis

During the sequence from which are extracted the stereo images above presented, the axial speed of the car have been measured. Results are presented on figure 7.

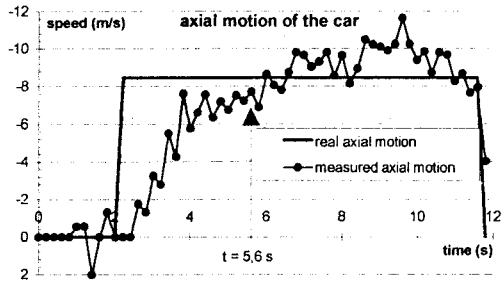

Figure 7. Evolution of motion for the car.

For constant motion, we can observe that the computed speed converges to the real one as expected by the process.

VI. Conclusion

We have presented an original system to determine axial motion maps by computing a range of disparity maps without spatio-temporal matching. The aim of our solution is in two steps. In the first step, we estimate axial motion with a low level computing, in the second step, we use a morphology operation to recover the structure. At present, 128X128 pixels stereo images are processed is less than 0.5 s on an IPX SUN workstation which is nearly real-time for some robotics applications. For other applications in car driving assistance, our system will be soon implemented on a specialised architecture comprising several DSP.

REFERENCES

[1] Peng S. L, Medioni G, "Interpretation of Image Sequences by Spatio-Temporal Analysis", IEEE conference, 344 - 351, 1989

[2] Matteucci P, Regazzoni C. S. and Foresti G. L, "Real-time approach to 3-D object tracking in complex scenes", ELECTRONICS LETTERS, VOL 30, N , 475 - 476, 1994

[3] Spetsakis M, Aloimonos J,"A Multi-frame Approach to Visual Motion Perception", International Journal of Computer Vision, VOL 6:3, 245 - 255, 1991

[4] Sabata B, Aggarwal J. K., "Estimation of Motion from a Pair of Range Images : A Review", CVGIP : IMAGE UNDERSTANDING, VOL 54, N° 3, 309-324, 1991

[5] Lin C. D, Goldgof D. B, Huang W, "Motion estimation from scaled orthographic projections without correspondences", Image and Vision Computing, VOL 12, N° 2, 95-108, 1994

[6] Li L, Duncan J.H, "3-D Translation Motion and Structure from Binocular Image Flows", IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL 15, N , 657-667, 1993

[7] Miché P, Bensrhair A, Lee S-G, "High Speed Self-Adaptive Algorithms for Implementation in a 3-D Vision Sensor", Journal of Korean Sensor Society. Vol 2, pp 123-130, 1997

[8] Oksenhenler V, Bensrhair A, Miché P, Lee S-G, "A 3-D Vision Sensor Implementation on Multiple DSPs TMS320C31", Journal of Korean Sensor Society. Vol 7, N° 2, pp 42-48, 1998

[9] Bensrhair A., Miché P., and Debrie R., "Fast and automatic stereovision matching algorithm based on dynamic programming method", Pattern Recognition Letter, N° 17, p. 457-466, 1996

著 者 紹 介

Stephane MOUSSET*

1963년 프랑스 출생. 1997년 프랑스 국립루앙대학교 졸업(공학박사), 현재 프랑스 국립 응용과학원(Institut National des Sciences Appliquees de Rouen), 정보 인식 시스템 / 센서 기기 및 분석 연구소(PSI/LCIA ; Perception Systeme Information / sensor instrumentation & analysis lab.) 연구원. 프랑스국립 루앙 대학교 물리측정학과 전임강사.

Abdelaziz BENSRHAIR*

『센서학회지 제7권 제2호』 논문 98-7-2-07 참조.
현재 프랑스 국립 응용과학원 (Institut National des Sciences Appliquees de Rouen), 정보 인식 시스템 / 센서 기기 및 분석 연구소 (PSI/LCIA ; Perception Systeme Information / sensor instrumentation & analysis lab.) 연구원. 프랑스국립 루앙 대학교 전자 전기공학부 조교수.

Pierre MICHE,

『센서학회지 제7권 제2호』 논문 98-7-2-07 참조.
현재 프랑스 국립 응용과학원(Institut National des Sciences Appliquees de Rouen), 정보 인식 시스템 / 센서 기기 및 분석 연구소(PSI/LCIA ; Perception Systeme Information / sensor instrumentation & analysis lab.)소장, 프랑스 국립 루앙대학교 공과대학 학장

Sang-Goog LEE

『센서학회지 제7권 제2호』 논문 98-7-2-07 참조.
현재 프랑스 국립 응용과학원 (Institut National des Sciences Appliquees de Rouen), 정보 인식 시스템 / 센서 기기 및 분석 연구소 (PSI/LCIA ; Perception Systeme Information / sensor instrumentation & analysis lab.)연구원, 경북대학교 센서기술연구소 프랑스현지연구센터 (IJRL)운영실장, 프랑스국립 루앙대학교 전자전기공학부 조교수