논문 97-6-2-06

3-D 비젼센서를 위한 고속 자동선택 알고리즘.

P.미세*, A.벤스하이르*, 이 상국 (李 相國)**

High Speed Self-Adaptive Algorithms for Implementation in a 3-D Vision Sensor

Pierre MICHE*, Abdelaziz BENSRHAIR*, Sang-Goog LEE**

요 약

이 논문은 다음과 같은 두가지 요소로 구성되는 독창적인 stereo vision system을 논술한다. declivity라는 새로운 개념을 도입한 자동선택 영상 분할처리 (self-adaptive image segmentation process) 와 자동선택 결정변수 (self-adaptive decision parameters) 를 응용하여 설계된 신속한 stereo matching algorithm. 현재, 실내 image의 depth map을 완성하는데 SUN-IPX 에서 3sec가 소요되나 연구중인 DSP Chip의 조합은 이 시간을 1초 이하로 단축시킬 수 있을 것이다.

Abstract

In this paper, we present an original stereo vision system which comprises two process: 1. An image segmentation algorithm based on new concept called declivity and using automaticic thresholds. 2. A new stereo matching algorithm based on an optimal path search. This path is obtained by dynamic programming method which uses the threshold values calculated during the segmentation process. At present, a complete depth map of indoor scene only needs about 3 s on a Sun workstation IPX, and this time will be reduced to a few tenth of second on a specialised architecture based on several DSPs which is currently under consideration.

KEY WORD : Stereo Vision System, Self-Adaptive, Dynamic Programming

Ⅰ. Introduction

3-D vision systems are essential for future intelligent robot since a robot works in a 3-D world. Among the different classic techniques, passive stereo vision is a very attractive approach especially because of its possibility of working in a large variety of conditions. The goal of stereo algorithms is to determine the three-dimenional distance, or depth, of objects from images taken from different viewpoints.

The computational stereo paradigm consists of three major steps: feature extraction, feature matching and three-dimensional reconstruction.

* LCIA, INSA de Rouen, Place E. Blondel BP 08
76131 Mont Saint Aignan Cedex France
tel:+(33) 02 35 52 84 07 fax:+(33) 02 35 52 84 83
e-mail : bensrhair@aigle.insa-rouen.fr
** Sensor Technology Research Center, Kyungpook Nat'l Univ.
<접수일자 : 1996년 11월4일>

Many stereo vision systems have been used in the past using different principles. However, most of these methods are computationally expensive or use fixed parameters. In our study, stereo vision is used as a basic principle in the development of 3-D vision sensor for applications in robots guidance and the navigation of autonomous vechicles. These applications needs fast, robust and self-adaptive algorithms to respond to a large variety of illumination.

We propose here a stereo vision system which comprises the following:

- an innovative fast data processing to extract feature points in the stereo images using self-adaptive decision parameters[1],
- a new fast stereo matching algorithm based on dynamic programming using self-adaptive parameters[3][4],

II. Feature extraction

To simplify the problem of camera calibration as well as feature extraction and matching, we choose a special stereo system that allows the stereo process to be quicker and the algorithms to be reduced to a one-dimensional problem.

In this configuration (figure 1), the cameras are set up so that their optical axes are parallel and their common baseline is parallel to the image horizontal lines[5].

Thus, let $P_L(X_L,Y_L)$ and $P_R(X_R,Y_R)$ be the two stereo corresponding points of an object point $P(X_P,Y_P,Z_P)$. Then, the relations used for the three-dimensional reconstruction are the following:

$$X_P=X_R.d/D$$
$$Y_P=Y_R.d/D$$
$$Z_P=f.d/D \qquad (1)$$

In these relations d is the length of the baseline, D is the horizontal disparity of the stereo corresponding points which is equal to X_R-X_L, and f is the focal length of the two cameras.

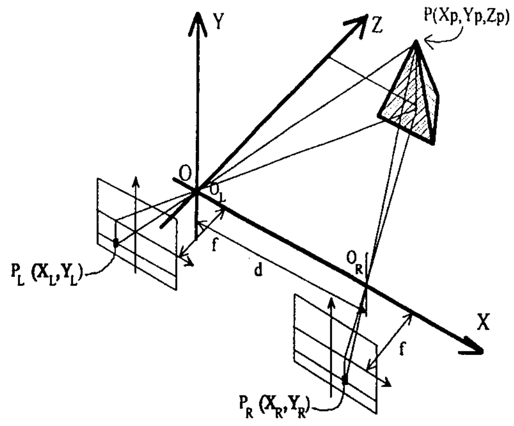


Figure 1. The stereo vision system configuration

With the stereo camera configuration presented above, our segmentation algorithm consists in extracting characteristic elements line by line. This is obtained by a self-adaptive thresholding of the declivities.

In an image line, a declivity is defined as a cluster of contiguous pixels, bounded by two end-points which correspond to two consecutive local extrema of the grey level intensity, i.e. one maximum and one minimum (see figure 2).

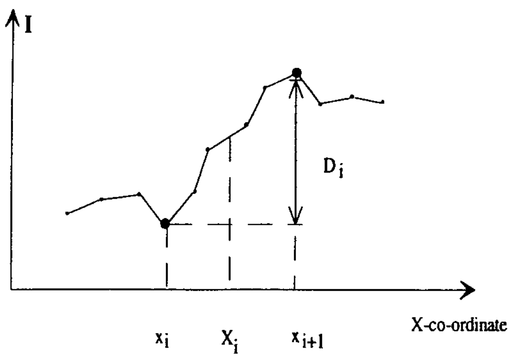


Figure 2. An example of a declivity

According to the definition of a declivity, an image line is equivalent to a set of declivities which can be used as basic elements for segmentation. For this purpose, we consider that declivities can be sorted by their amplitude into two classes. The first one contains low- amplitude declivities which correspond to noise and non-sinificant elements. The second class contains high-amplitude declivities. These last ones are used as features in our stereo vision process; in the following they are called characteristic declivities. It will be noticed that they correspond to edges.

From what previous, it appears that feature extraction will be obtained simply by thresholding the declivities amplitudes. For this operation to be self-adaptive, the threshold is computed from a statistical analysis of the image line. Finally, the threshold value d_t is defined by:

$$d_t = 5.6\sigma \tag{2}$$

where σ is the standard deviation of the noise which is considered as a white Gaussian one[1].

Once characteristic decclivities are detected, a major problem is the determination of their real positions. In effect, the accuracy of the 3-D component greatly depends on these positions. Finally, this problem was solved by computing the mean position of the declivity points weighted by the gradients squared as follows:

$$X_i = \frac{\sum_{x=x_i}^{x_{i+1}-1} [I(x+1) - I(x)]^2 (x+0.5)}{\sum_{x=x_i}^{x_{i+1}-1} [I(x+1) - I(x)]^2} \tag{3}$$

where X_i is the position of a declivity.

III. Declivity attributes

We recall that due to the special configuration of the two cameras, the application of the epipolar constraint to the matching problem reduces it to one dimension and implies that two corresponding image lines have the same number. Then, to be matched, declivities must satisfy geometric and photometric similarity which depend on attributes given below.

Let (R(i,l)) and (L(j,l)) be two sets of declivities ordered according to their co-ordinates in an arbitrary l right and l left epipolar lines, with i and j are the indexes of declivities. Each declivity is described by the following attributes:

- its X-coordinate on the image line defined by (3), which is used to calculate the disparity value,
- the grey levels of its three left neighbours pixels,
- and the grey levels of the three right neighbours pixels,

To estimate the photometric similarity between two declivities to be matched, we consider the left and right photometric distances defined by:

$$lpd = \sum_{k=0}^{2} |I_R(x_i - k) - I_L(x_j - k)| \tag{4}$$

$$rpd = \sum_{k=0}^{2} |I_R(x_{i+1} + k) - I_L(x_{j+1} + k)| \tag{5}$$

where $I_R(x_i-k)$, $I_R(x_{i+1}+k)$ are the grey levels of the kth left and kth right neighbours pixels of the declivity R(i,l) bounded by the extrema x_i and x_{i+1}, and $I_L(x_j-k)$, $I_L(x_{j+1}+k)$ are those of the declivity L(j,l).

IV. The matching algorithm

Our matching algorithm is based on dynamic programming. In this way, a 2-D graph is established whose rows and columns are affected to characteristic declivities of two corresponding lines. Intersections of rows and columns correspond to hypothetical declivity matching. Among them, those that meet the limitation of research area constraint define nodes on the graph. This geometric constraint is explained bellow.

The matching problem can be treated as the problem of finding an optimal path on this graph using the non-reversal constraint in primitives as well as the uniqueness constraint. To select this path, classic methods tend to minimise a cost function[6][7]. The main difficulty with this approach is that the cost value can increase indefinitely which affects the computation algorithm. In our approach, we chose to maximise a global gain.

The matching algorithm consists in three steps.

In the first step, a 2-D graph is established as above explained from the sets (R(i,l)) and (L(j,l)) of characteristic declivities of two corresponding lines. In this step, the limitation of research's area constraint is defined by:

$$0 < X_{Ri}-X_{Lj} < \text{disp max} \tag{6}$$

where X_{Ri} and X_{Lj} are the X-coordinates of the declivities R(i,l) and L(j,l) respectively , and dispmax is the maximum value of the disparity which has to be determined for each application environment.

In the second step, a local gain is allocated to each node of the graph. It is a non-linear function of the photometric distance expressed by:

$$\text{gain}=f(\text{gmax})-\text{lpd}-\text{rpd} \tag{7}$$

The term f(gmax) has a double object. First it ensures the self-adaptivity of the gain by gmax which is defined by:

$$\text{gmax}=0.5(d_{tR}+d_{tL}) \tag{8}$$

where d_{tR} and d_{tL} are the right and left self-adaptive threshold values defined by (2) respectively. Then, it enables to take account partially masked objects which induces three case of similarity: right and left, only right or only left. Thus the following formulas are used:

- If lpd < gmax and rpd < gmax, then f(gmax) is initialised to 3 gmax.
- If lpd < gmax or rpd < gmax, then f(gmax) is initialised to gmax.
- If lpd > gmax and rpd > gmax, then the corresponding hypotheses are rejected.

The gain of the path (global gain) is the sum of those of its primitive paths. The gain is defined as follows: let G(m,n) be the maximum gain of the partial path from an initial node - not necessarily (1,1) - to node (m,n) as shown in figure 3, and let g(m,n,i,j) be the gain corresponding to the primitive path from node (m,n) to node (i,j), with l<=m<i and l<=n<j, and which, in fact, only depends on the node (i,j) as defined by (7).

Finally, G(i,j) is given by:

$$G(i,j) = Max_{(m,n)}[G(m,n)+g(m,n,i,j)] \tag{9}$$

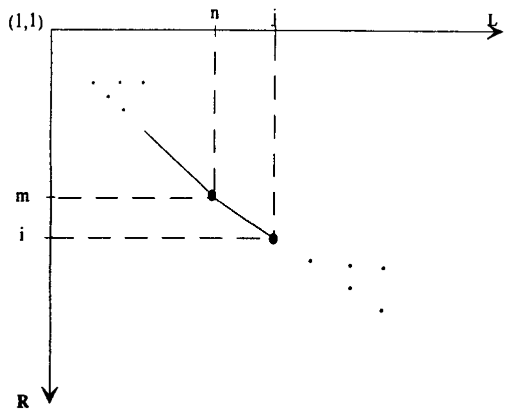


Figure 3. The matching graph elements

As a result, we obtained several paths from the origin node to the final node in the graph (an experimental example is given on figure (4).

In the final step, the path with the maximum global gain is selected which gives the matching declivities. From this, we can thus deduct a disparity value: disp=$X_{Ri}-X_{Lj}$, then, every declivity in the right epipolar line is assigned to its disparity value. For points which do not correspond to declivities, the disparity is assigned by interpolation. A simple interpolation is done between declivities positions where the disparity is obtained. Finally, finding the 3-D position of the object is obtained from (1).

V. Experimental results

Some experiments were performed on an IPX SUN workstation (running UNIX). Digitised images have a size of 512 x 512 with an intensity resolution of 256. Our algorithms have been implemented in C language and tested on indoor scene images as well as on outdoor ones. In this paper, we only present results concerning an indoor scene.

Experimental parameters are: f=8 mm for the local length of the two cameras, d=8 cm for the base line length and dispmax=60.

Results of the declivities detection by our segmentation algorithm are shown in photo 3 and 4. The complete run-time for declivity extraction in a line of the indoor scene (photos 1 and 2) is only 3.1 ms on an IPX, and 2.3 ms on a DSP TMS 320C30 from Texas Instruments. The positions of characteristic declivities are known with sub-pixel accuracy.

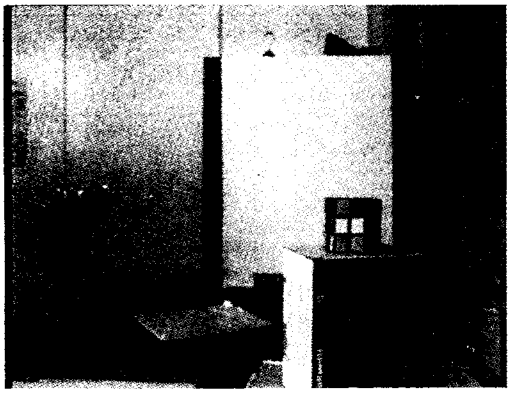


Photo 1. Left stereo image.

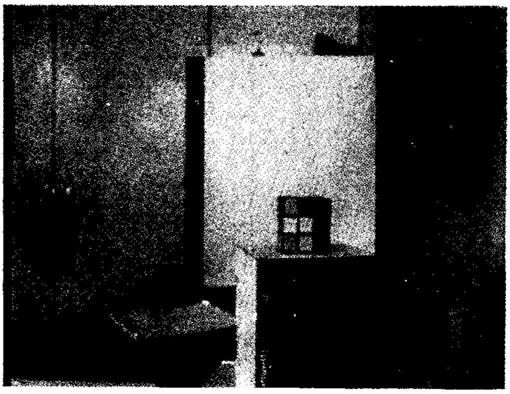


Photo 2. Right stereo image.

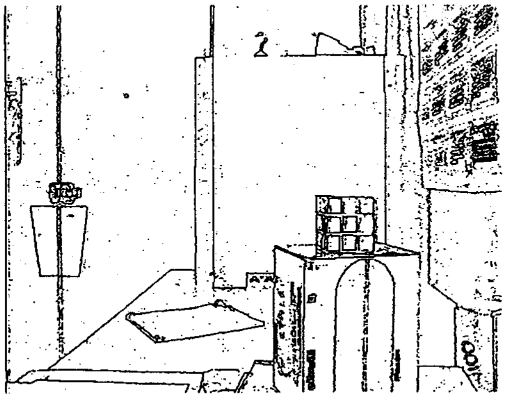


Photo 3. Left deckivities detection.

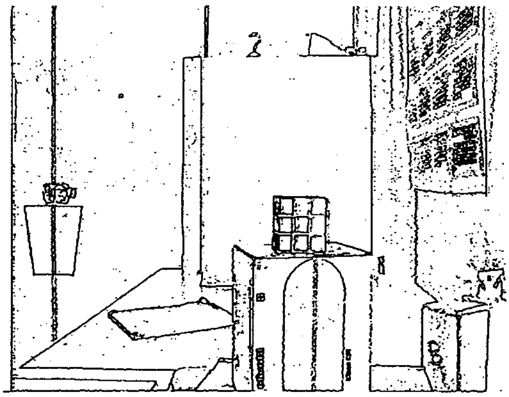


Photo 4. Right declivities detection.

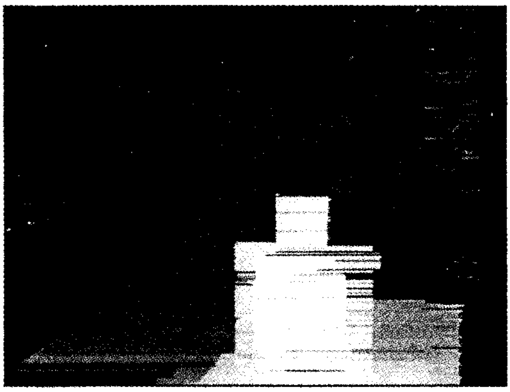


Photo 5. Depth map.

R/L	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	$\boxed{153}$			30																					
2					$\boxed{151}$		37																		
3						$\boxed{114}$																			
4								$\boxed{158}$																	
5									$\boxed{149}$																
6										$\boxed{129}$			38												
7											$\boxed{131}$	19			39										
8											21	$\boxed{152}$				16									
9													$\boxed{125}$	19			35								
10															90	105									
11													46	$\boxed{124}$			12	35							
12															$\boxed{145}$				137	23					
13																$\boxed{141}$				31		33			
14																	$\boxed{133}$				29		43		
15																	19	$\boxed{32}$			47		23		132
16																38			32	$\boxed{137}$		49			
17																	23	45			$\boxed{140}$				44
18																				34	97		$\boxed{123}$		
19																						98		$\boxed{43}$	
20																									$\boxed{137}$
21																							24		

Fig. 4. An example of an experimental matching graph

Concerning the matching algorithm, we obtained very satisfactory results: we only need 3.0 ms to match two image lines with about 30 declivities or 2.2 ms on a DSP, and, on average, more than 98% of declivities appear to be correctly matched.

An example of an experimental matching graph is given on figure 4; the vertical axis represents indexes of declivities in the right scanline and the horizontal one represents those in the corresponding left scanline. In this graph, the optimal path is shown: it is constituted by double edged nodes.

Once applied the matching algorithm to the whole image, we obtain a depth map (photo 5) which is coded using grey level value as follows: high grey level points are closer to the viewer.

The computation time for a complete processing of a depth map is only about 3 s on the IPX workstation and 2.3 s on DSP.

Using these depth maps, we deduct computed distances (Zp) of some objeccts in the scene. The result of computed distances is satisfactory. Indeed, on average, the errors are of the order of 4% in Z of around 1m and 11% in Z of around 4m.

Ⅵ. Conclusion

The purpose of our study is the realisation of a 3-D vision sensor which can be used in the field of the navigation of robots and autonomous vehicles.

These applications require fast, self-adaptive algorithms which can be processed by parallel processors. This was obtained by means of a special configuration and a highly parallelisable stereo vision process based on the declivity feature matched by dynamic programming.

To get benefit of the parallelisable characteristic of our algorithms, we are currently implementing our method on machine vision which comprises several DSPs thanks to which we expect a complete depth map computation time to be reduced to few tenths of second.

Recently, an axial motion process of dynamic scenes have been presented[8], it is based on the technique presented in this paper.

Ⅶ. References

[1] Miché P., and Debrie R.,. Fast and self-adaptive image segmentation using extended declivity, Annals of telecommunication, 50, No 3-4, 401-410, 1995.

[2] Miché P. Bensrhair A. and Debrie R. "Fast and self-adaptive stereovision algorithms for implementation in a 3-D vision sensor" Proceedings of the 3rd IEEE Internationa Conference on Advanced Motion Control AMC (1994), p.993-1001, BERKELEY, USA.

[3] Bensrhair A., Miché P., and Debrie R. "Fast and automatic stereo vision matching algorithm based on dynamic programming method. Pattern Recognition Letters Vol.17, No.5, May 1996

[4] Bensrhair A., Miché P., and Debrie R.. "New fast stereo matching for parallel processing" Proceeding of the 2nd IEEE International Worhshop on Advanced Motion Control AMC, p.87-92, Nagoya, Japan, 1992.

[5] Xiao-Wei T. and Dubuisson B. "3-D information erivation from a pair of binocular images" Pattern Recognition, No.23, 3/4, p.223-235, 1990.

[6] Baker H.H., and Binford T.O. "Depth from edge and intensity based stereo" Proc.7th Int. Joint Conference on Artificial Intelligence, p.631-636, Vancouver, Canada, 1981.

[7] Otha Y. and Kanade T. "stereo by intra- and inter-scanline search using dynamic programming" IEEE PAMI, 7(2), p.139-154, 1985.

[8] Mousset S., Miché P. Bensrhair A. and Debrie R. "Real-time determination of axial motion component from a stereo image sequence, proc. 5th Int. Conf. on Image Processing and its Applications, Edinburgh, U.K. p.246-250, 1995.

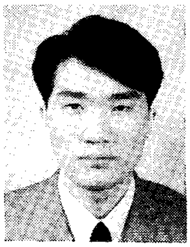
著 者 紹 介

Pierre MICHE

1947년 프랑스 생. 1975년 프랑스 국립 루앙 대학교 졸업(공학박사). 프랑스 루앙 국립 기술대학교 (Institut Universitaire de Technologie de Rouen)학장. 프랑스 국립 응용 과학원(Institut National des Sciences Appliquées de Rouen) 의 L.C.I.A. (Sensor, Instrumentation and Analysis Laboratory) 에서 vision system 연구팀장.

Abdelaziz BENSRHAIR

1960년 2월 15일 모로코 생. 1992년 루앙대학교 전자.전기공학부 computer vision 전공 졸업 (공학박사). 프랑스 루앙 국립 기술대학교(I.U.T. de Rouen)전자.전기공학부 조교수. 프랑스 국립 응용 과학원(L.C.I.A., INSA de Rouen) vision system 연구팀 연구원.



Sang-Goog LEE

1962년 12월 23일 생. 1988년 인하대학교 전기공학과 졸업(학사). 1990년 인하대학원 졸업(공학석사). 1994년 프랑스 Université de Rouen 전자.전기공학부 졸업(공학박사). 1990년- 프랑스 국립 응용 과학원(L.C.I.A., INSA de Rouen)연구원, 1994- 경북대학교 센서기술연구소 선임연구원, 프랑스 루앙 국립 기술대학교 (I.U.T. de Rouen) 전자.전기공학부 조교수.