A FAST MULTILEVEL METHOD FOR MATCHING STEREO IMAGES

M. Hariti, Y. Ruichek and A. Koukam

Laboratoire Systèmes et Transports
Université de Technologie de Belfort-Montbéliard
90010 Belfort Cedex France

Abstract? This paper presents a method, which performs edge stereo matching at different levels, from significant edges to less significant ones. At each level, the process starts by selecting significant edges with respect to their gradient magnitude. The selected edges are then matched in order to obtain reference edges from which less significant edges will be matched in the next level. The matching procedure is based on a voting scheme by using local and global constraints. The performance of the proposed stereo matching method is evaluated for real-time obstacle detection in front of a vehicle using linear cameras.

I. INTRODUCTION

Passive stereo vision is a well known technique for obtaining 3-D depth information of objects seen by two or more video cameras from different viewpoints [1][2]. The key problem in this approach is the matching process which is difficult to solve and computationally expensive [3]. It consists in comparing each feature extracted from one image with a number, generally large, of features extracted from the other image in order to find the corresponding one, if any. Once the matching process is established and the stereo vision system parameters are known, the depth computation is reduced to a simple triangulation technique.

Many authors circumvent the stereo matching problem by making hypotheses about the type of objects being observed and their visual environment [4]. With such restrictive assumptions, the number of candidate features for matching is substantially reduced so that the computing times become acceptable for real-time applications without an important loss of useful information. Unfortunately, setting-up a conventional stereo vision system on board a moving vehicle for real-time obstacle detection is difficult because, in the road environment, the features are too numerous to allow a reliable matching within an acceptable computer time [5].

Considering these difficulties, some authors have proposed to use linear cameras instead of matrix ones [5-7]. With these cameras, the information to be processed is drastically reduced since their sensor contains only one video line, typically 2,500 pixels, instead of 250,000 pixels with standard raster-scan cameras. Furthermore, they have a better horizontal resolution than video cameras. This characteristic is very important for an accurate perception of the scene in front of a vehicle.

The aim of this research is to solve the stereo correspondence problem in order to provide reliable information for detecting and localizing obstacles in front of a moving car, using a linear stereo vision set-up. This problem has been previously solved using correlation methods [5][8]. With these methods, a majority of candidate features can be matched without ambiguities. However, these sequential methods can leave some unmatched edges, and may lead to false matches, which are difficult to identify. Other authors have proposed to turn the matching problem into an optimization task where an objective function, representing the constraints on the solution, is to be minimized using a Hopfield neural network or genetic algorithms [9][10]. These optimization approaches provide good results when compared to the former ones using correlation. Their major disadvantage is the computation time, which is incompatible with real-time obstacle detection in front of a moving vehicle.

In this paper, we propose a fast multilevel stereo matching method for real-time obstacle detection. The method performs edge stereo matching at different levels, from significant edges to less significant ones. At each level, the process starts by selecting significant edges with respect to their gradient magnitude. The selected edges are then matched and the obtained pairs are used as reference pairs for matching less significant edges in the next level. The matching procedure is performed thanks to a voting scheme by using local and global constraints. Local constraints are used to discard impossible matches whereas global ones are used to implement the voting strategy.

The remainder of this paper is organized as follows. Section 2 presents the basic principles of stereo vision with linear cameras. The method used to extract edges from linear images is described in section 3. Section 4 presents the multilevel stereo matching method. Before concluding, experimental results for real-time obstacle detection are presented in Section 5.

II. LINEAR STEREO VISION

A linear stereo system is build with two line-scan cameras, so that their optical axes are parallel and separated by a distance E (Fig. 1). Their lenses have a same focal length f. The fields of view of the two cameras are merged in the same plane, called optical plane, so that the cameras shoot the same scene [11]. If any object intersects the stereo vision sector, which is the common part of the two fields of vision in the optical plane, it produces a disparity between the two stereo linear images and, as a consequence, can be localized by means of triangulation technique.

Let us define the base-line joining the perspective centers O_l and O_r as the X-axis, and let Z-axis lie in the optical plane, parallel to the optical axes of the cameras, so that the origin of the $\{X,Z\}$ coordinate system stands midway between the lens centers (Fig. 2). Let us consider a point $P(x_p,z_p)$ of coordinate x_p and z_p in the optical plane. The image coordinates x_l and x_r represent the projections of the point P in the left and right imaging sensors, respectively. This pair of points is referred to as a corresponding pair.

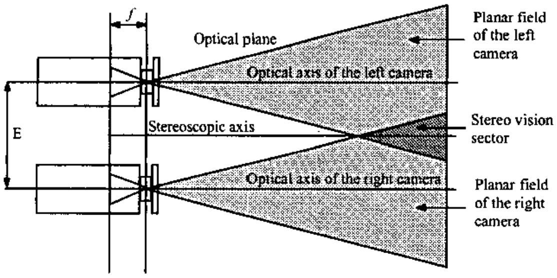

Fig. 1. Geometry of the linear cameras.

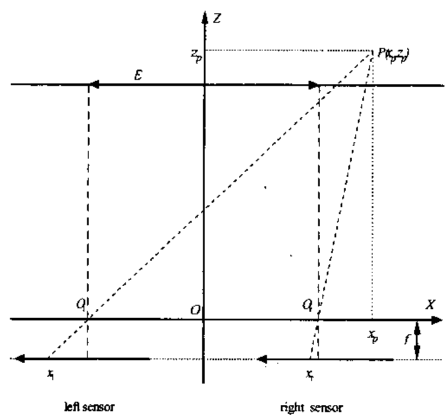

Fig. 2. Pinhole lens model.

Using the pinhole lens model, the coordinates of the point P in the optical plane can be found as follows:

$$z_p = \frac{E \cdot f}{d}, \tag{1}$$

$$x_p = \frac{x_l \cdot z_p}{f} - \frac{E}{2} = \frac{x_r \cdot z_p}{f} + \frac{E}{2}, \tag{2}$$

where f is the focal length of the lenses, E is the base-line width and $d = |x_l - x_r|$ is the disparity between the left and right projections of the point P on the two sensors.

III. EDGE EXTRACTION

Edge extraction is performed by means of the Deriche's operator [12]. After derivation, a procedure is applied to select pertinent local extrema. This is achieved by splitting the gradient magnitude signal into adjacent intervals where the sign of the response remains constant (Fig. 3). In each interval of constant sign, the maximum amplitude indicates the position of a unique edge associated to this interval when, and only when, this amplitude is greater than a low threshold value t. The application of this threshold value allows to remove the very small responses of the differential operator lying in the range $[-t,+t]$. The adjustment of t is not crucial. Good results have been obtained with t adjusted at *10%* of the greatest amplitude of the response of the differential operator.

Applied to the left and right linear images, this edge extraction procedure yields two lists of edges. Each edge is characterized by its position in the image, the amplitude and the sign of the response of Deriche's operator.

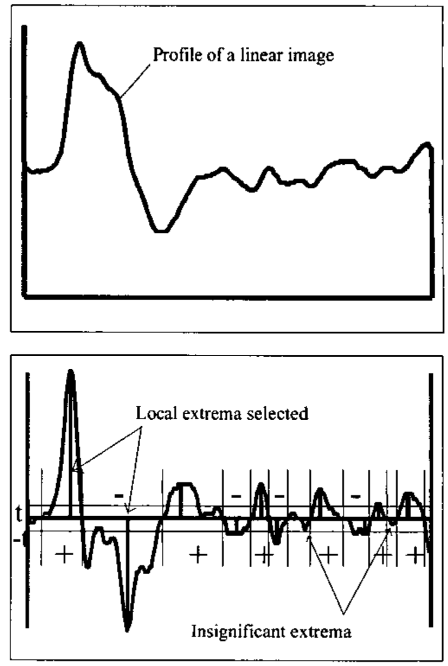

Fig. 3. Edge extraction.

IV. EDGE STEREO MATCHING

A) Multilevel stereo matching scheme

The stereo matching between the edges extracted from the left and right images is performed by a multilevel scheme. This is achieved by matching the edges at different levels, from the significant edges to less significant ones, with respect to their gradient magnitude. At each level, and from a current research zone, the matched edges define new research zones for the matching process in the next level. At the first level, the research zone is defined by the two stereo images, i.e., by considering all the edges in the left and right images (Fig. 4). For a given level n, and from a current research zone, the process starts by selecting significant edges for which the gradient magnitude satisfies the following relations:

for edges in the left image:

$$\frac{\min_l}{2^{n-1}} \le mg_l \le \frac{\min_l}{2^n} \; or \; \frac{\max_l}{2^n} \le mg_l \le \frac{\max_l}{2^{n-1}} \tag{3}$$

for edges in the right image:

$$\frac{\min_r}{2^{n-1}} \le mg_r \le \frac{\min_r}{2^n} \; or \; \frac{\max_r}{2^n} \le mg_r \le \frac{\max_r}{2^{n-1}} \tag{4}$$

where min_l and max_l (respectively min_r and max_r) are the smallest and biggest gradient magnitudes of the edges extracted from the left (respectively right) image. mg_l and mg_r are the gradient magnitudes of the edges l and r extracted from the left and right images, respectively.

The selected edges are then matched using a voting procedure, which is described in the next section. The obtained pairs, called reference pairs, define new research zones in which the same process is applied to match less significant edges in the next level (Fig. 4). In order to optimize the running of the multilevel scheme, the procedure is implemented recursively.

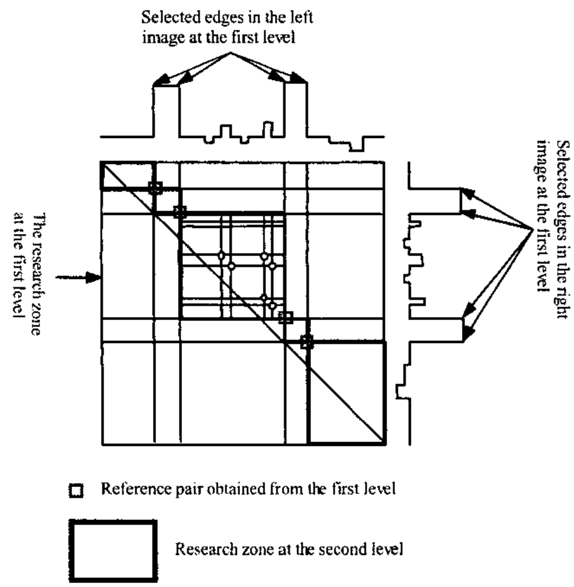

Fig. 4. stereo matching multilevel scheme.

B) Stereo matching voting method

For a given level and for a given research zone, let L and R be the lists of the edges selected from the left and right images, respectively, by using equations (3) and (4). The stereo matching problem is first mapped onto a $N_L x N_R$ matrix M, called matching matrix, where N_L and N_R are the numbers of edges in L and R, respectively. Each element M_{lr} of the matrix explores the hypothesis that the edge l in the left image matches the edge r in the right one. We consider only valid elements which represents possible matches with respect to two local constraints (Fig. 5). The first one is a geometric constraint, which assumes that the edges l and r represent a possible match only if the constraint $x_l > x_r$ is satisfied, where x denotes the position of the edge in the image. The second local constraint is the slope constraint, which means that only edges with the same sign of the gradient are considered for a possible matching.

After the mapping step, the stereo matching process is performed thanks to a voting procedure. Each element M_{lr} of the matrix is characterized by a score, which can be interpreted as a degree of compatibility of the pair *(l,r)* with respect to correct matches. The voting strategy is based on three global constraints. The first one is the uniqueness constraint, which assumes that one edge in the left image matches only one edge in the right image (and vice versa). The second global constraint is the ordering constraint, which means that if an edge l in the left image is matched with an edge r in the right image, then it is impossible for an edge l' in the left image, such that $x_{l'} < x_l$, to be matched with an edge r' in the right image, for which $x_{r'} > x_r$. The third global constraint is the smoothness constraint, which assumes that neighboring edges have similar disparities.

For each element M_{lr}, the voting strategy consists first of determining the elements, which will vote for the candidate M_{lr}. Those voting elements are determined by using the uniqueness and ordering constraints (Fig. 5): an element $M_{l'r'}$ will vote for a candidate M_{lr} if the pairs *(l,r)* and *(l',r')* verify the uniqueness and ordering constraints. Each voting element $M_{l'r'}$ performs than its vote by contributing to the score of the candidate M_{lr}. This voting contribution, which uses the smoothness constraint, is computed as follows:

$$SM_{l'r'}(new) = SM_{l'r'}(previous) + f(X_{lrl'r'}) \tag{5}$$

where $X_{lrl'r'}$ is the absolute value of the difference of disparities of the possible matches *(l,r)* and *(l',r')*, expressed in pixels, and f is a nonlinear function given by:

$$f(X) = \frac{1}{1+X} \tag{6}$$

Once the voting process is achieved, and to determine the pairs of corresponding edges, a procedure is designed to select the elements for which the score is maximum in each row and column of the matching matrix.

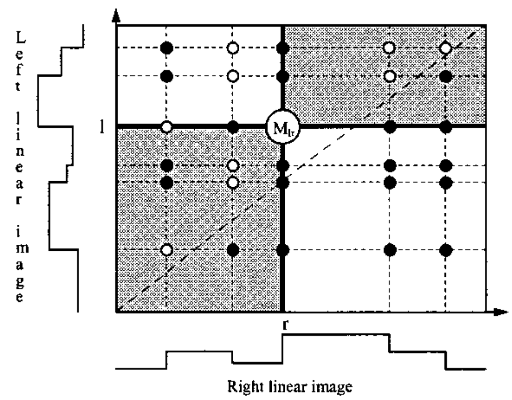

Fig. 5. Matching matrix. The white circles represent the possible matches whereas the black ones represent the impossible matches with respect to the local constraints. The elements lying in the gray area represents the voting elements for the candidate M_{lr}.

V. APPLICATION TO OBSTACLE DETECTION

The performance of the proposed matching method is evaluated for real-time obstacle detection in front of a vehicle using a linear stereoscopic sensor. A stereo set-up, built with two line-scan cameras, is installed on top of a car for periodically acquiring stereo pairs of linear images as the car travels (Fig. 6). The tilt angle is adjusted so that the optical plane intersects the pavement at a given distance D_{max} in front of the car.

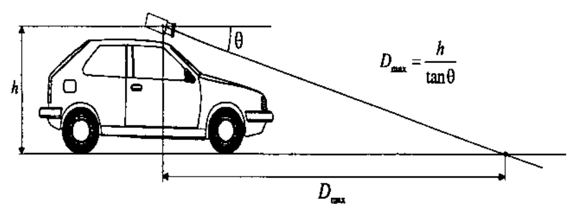

Fig. 6. Geometry of the stereo set-up.

One of the sequences shot by the linear stereo set-up is shown in Fig. 7, where the linear images are represented as horizontal lines, time running from top to bottom. In this sequence, the prototype car travels in the central lane of the road and follows another car. The optical plane intersects gradually the shadow of the preceding car, then the whole car from the bottom to the top, as the prototype car comes near to it. A third car pulls back into the central lane after overtaking the preceding car. The prototype car is itself overtaken by another car which is traveling in the third lane of the road.

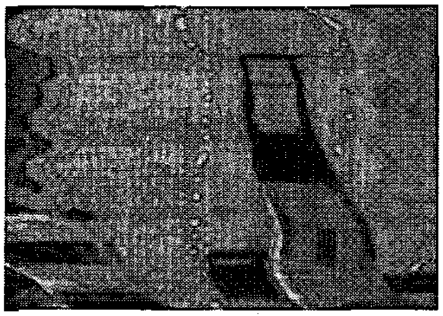

Fig. 7. Left sequence acquired by the set-up.

We can see in the pictures of Fig. 7 the white lines which delimit the pavement of the road and, between these lines, the two dashed white lines and the preceding car. At the bottom of the pictures, i.e., at the end of the sequence, we can see on the left most lane, the car which is overtaking the prototype car and, in the middle, the shadow of the vehicle which pulls back in front of the preceding car. The curvilinear aspect of the lines is due to the changes in the stereoscope tilt because of the uneven road surface. Note that the depth reconstruction is not affected by these oscillations of the car, provided the optical planes of the two cameras remain correctly calibrated when the car is running.

This stereo sequence has been processed using the proposed multilevel matching algorithm. The disparities of all matched edges are used to compute the positions and distances of the edges of the objects seen in the stereo vision sector. The results are shown in Fig. 8, in which the distances are represented as gray levels, the darker to closer, whereas positions are represented along the horizontal axis. As in Fig. 7, time runs from top to bottom.

Fig. 8. Reconstructed scene.

The proposed method provides good matching results. The edges of the two dashed lines have been correctly matched. The edges of the lines which delimit the road cannot be matched continuously because they do not always appear in the common part of the fields of the cameras. The preceding vehicle is well detected as it comes closer and closer to the prototype car as time runs. The shadow of the vehicle, which pulls back in front of the preceding vehicle, is identified as a white continuous line, at the bottom of the reconstructed image. Finally, at the bottom of the reconstructed image, we can see the dark oblique line, which represents the vehicle overtaking the prototype car.

The processing is performed with a *PC Intel-Pentium III* running at *1 GHz*. The time processing of the stereo sequence, which is composed by *200* stereo linear images, is *50 ms*. The average processing rate is hence *4000* pairs of stereo linear images per second.

VI. CONCLUSION

A fast multilevel method for matching stereo images is presented. The method performs edge stereo matching at different levels, from significant edges to less significant ones. At each level, and from a current research zone, the process starts by selecting significant edges with respect to their gradient magnitude. The selected edges are then matched and the obtained pairs define new research zones in which the same process is applied for matching less significant edges in the next level. The matching procedure is based on a voting strategy by using local and global constraints. The performance of the proposed stereo matching method is evaluated for real-time obstacle detection using linear cameras. Experimental results show that the method provides good depth computation with a very interesting processing rate.

REFERENCES

[1] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs: Prentice Hall, 1982.

[2] D. Marr, and T. Poggio, "Cooperative Computation of stereo Disparity," *Science*, vol. 194, pp. 283–287, 1976.

[3] S. Barnard, and M. Fisher, "Computational stereo," *ACM Comput. Surveys*, vol. 14, pp. 533–572, 1982.

[4] D. J. Kriegman, E. Triendl, and T. O. Binford, "Stereo vision and navigation in buildings for mobile robots," *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, 1989.

[5] J. L. Bruyelle and J. G. Postaire, "Direct range measurement by linear stereo vision for real-time obstacle detection in road traffic," *Robotics and Autonomous Systems*, no. 11, pp. 261–268, 1993.

[6] R. M. Inigo and T. Tkacik, "Mobile robot operation in real-time with linear image array based vision," *in Proc. IEEE Int. Intelligent Control Symposium*, Philadelphia, 1987, pp. 228–233.

[7] O. Colle, "Vision stéréoscopique à l'aide de deux caméras linéaires - Application à la robotique mobile," Ph.D. dissertation, National Institute of Applied Sciences, Lyon, France.

[8] J. C. Burie, J. L. Bruyelle, and J. G. Postaire, "Detecting and localising obstacles in front of a moving vehicle using linear stereo vision," *Mathematical and Computer Modelling*, vol. 22, no. 4–7, pp. 235–246, 1995.

[9] Y. Ruichek and J. G. Postaire, "A neural matching algorithm for 3-D reconstruction from stereo pairs of linear images," *Pattern Recognition Letters*, no. 17, pp. 387–398, 1996.

[10] Y. Ruichek, H. Issa, and J.G. Postaire, "Genetic approach for obstacle detection using linear stereo vision," *in Proc. IEEE Intelligent Vehicles Symposium*, Detroit, 2000, pp. 261–266.

[11] J. L. Bruyelle, "Conception et réalisation d'un dispositif de prise de vue stéréoscopique linéaire – Application à la détection d'obstacles à l'avant des véhicules routiers," Ph.D. dissertation, University of Sciences and Technologies of Lille, France, 1995.

[12] R. Deriche, "Fast algorithms for low-level vision," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 12, no. 1, pp. 78–87, 1990.