A Voting Stereo Matching Method for Real-Time Obstacle Detection

Mohamed Hariti, Yassine Ruichek, Abderrafiaa Koukam
Laboratoire Systèmes et Transports
Université de Technologie de Belfort-Montbéliard
90010 Belfort Cedex France
{mohamed.hariti,yassine.ruichek,abder.koukam}@utbm.fr

Abstract - **Depth from stereo is one of the most active research areas in the computer vision field. The heavily investigated problem in stereo approaches is the matching between two or more images of a scene observed, by two or more video cameras, from different viewpoints. It consists of identifying features in the left and right images that are projections of the same physical feature in the three-dimensional world. This paper presents a real-time stereo matching method using a voting schema. The correspondence problem is first mapped onto a two-dimensional matrix, called matching matrix, where each element represents a possible match between two features extracted from the left and right images. Local and global constraints are then used to search the true matches. The voting process is performed between the valid elements of the matching matrix, which represent compatible matches. The valid elements are determined by applying the local constraints. Global constraints are used to define the voting rules between the valid elements. The voting based-method is evaluated for real-time obstacle detection in front of a moving car using linear stereo vision.**

1. 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 [1-4]. The key problem in this approach is the matching process, which is difficult to solve and computationally expensive [5]. 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.

In the robot vision domain, this problem is generally simplified by making hypotheses about the type of objects being observed and their visual environment so that structural features, such as corners or vertical straight lines, can be more or less easily extracted [6]. With such restrictive assumptions, the number of candidate features for matching is substantially reduced so that computing times become acceptable for real-time processing without an important loss of useful information. Unfortunately, none of these hypotheses can be used in the road environment for detecting and localizing obstacles in front of a moving vehicle [7].

Considering these difficulties, some authors have proposed to use linear cameras instead of matrix ones [7-9]. 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.

In this paper, we propose a voting based-method to achieve real-time stereo matching for obstacle detection in front of a moving car using linear stereo vision. The correspondence problem is first mapped onto a two dimensional matrix, called matching matrix, where each element represents a possible match between two features extracted from the left and right images. The problem is then formulated in terms of finding true matches that satisfy local and global constraints [10]. Local constraints, named geometric and slope constraints, are used to determine, from the matching matrix, the valid elements representing the compatible matches. In the voting process, each valid element is considered at the same time as a voting element and candidate element. Each valid element is characterized by a matching voting score. Global constraints, named uniqueness, ordering and smoothness constraints, are used to define the rules in the voting process. The voting process uses two rules. The first one determines for each voting element the candidate elements concerned by its vote. This is performed by applying the uniqueness and ordering constraints. The second rule, which uses the smoothness constraint, consists of updating the voting scores of the candidate elements. After considering all the voting elements, and for determining the best matches, a procedure is designed to select the elements for which the voting scores are maximum in rows and columns of the matching matrix.

To demonstrate its effectiveness, the proposed stereo matching method is evaluated with experimental results for real-time obstacle detection using linear stereo vision.

The remainder of this paper is organized as follows. The following section describes the method used to extract prominent features from linear images. Section 3 presents the voting stereo matching method. Before concluding, experimental results for real-time obstacle detection in front of a moving car are presented in section 4.

II. FEATURE EXTRACTION

The first step in stereo vision consists in extracting significant features, representing physical objects, from the left and right images. Edges appearing in linear images, which are unidimensional signals, are valuable candidates for matching because large local variations in the gray-level function correspond to the boundaries of objects being observed in a scene.

Edge detection is performed by means of the recursive differential operator proposed by Deriche [11]. Before derivation, each linear image is first processed with a recursive smoothing filter, which removes noise while preserving edges. The gradient magnitude image indicates the amplitude and the sign of the derivative of the smoothed signal.

A very popular procedure for extracting significant edges consists in selecting the gradient magnitude extrema, which are greater than a given threshold. However, in practical situations, it is difficult, or even impossible, to adjust the threshold value for selecting all the significant edges while eliminating the irrelevant ones.

To overcome this problem, we propose first to use a low threshold value t, which is only used 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. A procedure is then applied to select the pertinent local extrema among the remaining edges. This is achieved by splitting the gradient magnitude signal into adjacent intervals where the sign of the response remains constant (Fig. 1). 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 t.

This edge extraction method is applied to the left and right linear images and yields two lists of edges as outputs. Each edge is characterized by its position in the image, the amplitude and the sign of the response of Deriche's operator. These two lists are the input of the matching procedure, which is described in the next section.

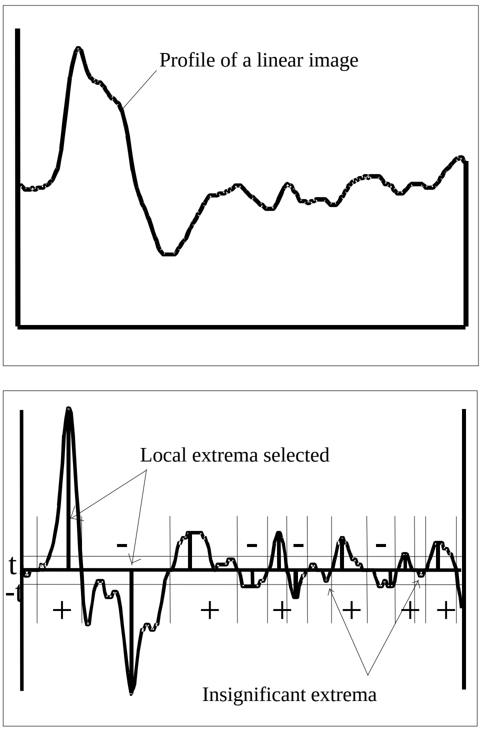


Fig.1 Edge extraction

III. THE VOTING STEREO MATCHING METHOD

The stereo matching problem is first mapped onto a $N_L \times N_R$ matrix, called matching matrix, where N_L and N_R are the numbers of edges extracted from the left and right images, respectively. An element M_{lr} of the matching matrix explores the hypothesis that the edge l in the left image matches the edge r in the right one.

In the matching matrix, we consider only valid elements, which represent compatible matches with respect to two local constraints. Resulting from the sensor geometry, the first one is a geometric constraint, which assumes that the edges l and r appearing in the left and right images, respectively, 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 to be matched. Fig. 2 shows an example of the matching matrix after applying the local constraints, where white circles represent the compatible matches whereas black ones represent the incompatible matches.

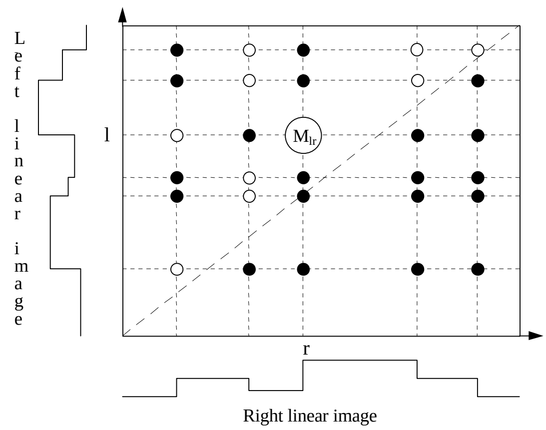


Fig.2 Matching matrix

The valid elements of the matching matrix are considered at the same time as voting elements and candidate elements in the voting process. Each valid element M_{lr} is characterized by a matching voting score VSM_{lr}, which represents an evaluation of the matching between the features l and r.

After the mapping step, the voting process is then performed thanks to 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 tends to preserve the order between the matched edges. This 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 $x_{r'} > x_r$. Note that this constraint can be violated in some cases, which are seldom encountered in real traffic conditions, since the base-line of the stereoscopic sensor is very small with respect to the size of the moving objects and that of the scene. The third global constraint is the smoothness constraint, which assumes that neighboring edges have similar disparities.

The voting process is governed by two rules. The first one determines for each voting element the candidate elements concerned by its vote. These candidate elements are determined by applying the uniqueness and ordering constraints (Fig. 3): a voting element M_{lr} votes for a candidate element $M_{l'r'}$ if the pairs *(l,r)* and *(l',r')* verify the uniqueness and ordering constraints. As illustrated in Fig. 3, the candidate elements for which the voting element M_{lr} will vote are the valid elements lying in the gray area.

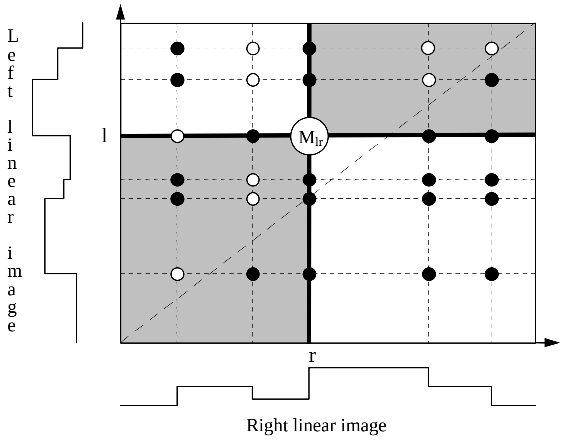


Fig.3 First rule of the voting process

The second rule of the voting process, which uses the smoothness constraint, consists of updating the voting scores of the candidate elements. Let M_{lr} be a voting element for a candidate element $M_{l'r'}$. The voting score $VSM_{l'r'}$ of $M_{l'r'}$ is updated as follows:

$$VSM_{l'r'}(new) = VSM_{l'r'}(previous) + f(X_{lrl'r'}) \quad (1)$$

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

$$f(X) = \frac{1}{1+X} \quad (2)$$

Note that the voting scores of the valid elements are initialized to *0*.

When all the voting elements have voted, the winner elements, which represent the correct matches, are those for which the voting scores are maximum.

Once the voting process is achieved, and to determine the matched edges, a procedure is designed to select the elements for which the voting scores are maximum. To eliminate ambiguities of multiple matches, the selection procedure is applied to rows and columns of the matching matrix.

IV. APPLICATION TO OBSTACLE DETECTION

The performance of the proposed stereo matching method is evaluated for obstacle detection in front of a moving car using linear stereo vision.

A stereo set-up is built with two line-scan cameras. Their lenses have a same focal length *f*. Their parallel optical axes, which define an optical plane, are separated by a distance *E* (Fig. 4).

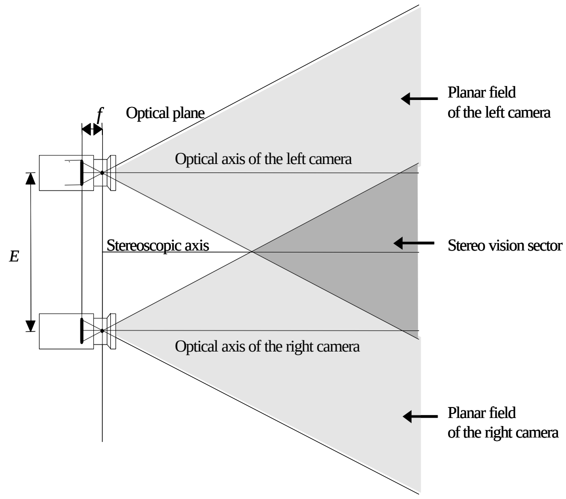


Fig.4 Geometry of the linear cameras

The stereo set-up is installed on top of a car for periodically acquiring stereo pairs of linear images as the car travels (Fig. 5). The tilt angle is adjusted so that the optical plane intersects the pavement at a given distance D_{max} in front of the car. This configuration ensures that every object that lies on the road in front of the vehicle is seen by the two cameras, even if its height is very small.

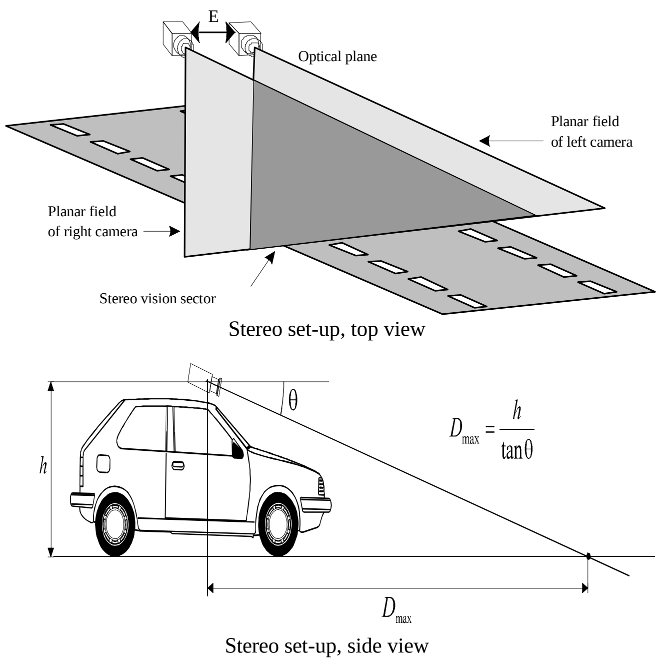


Fig.5 Geometry of the stereo set-up

Note that the fields of view of the two cameras must be merged in the same plane so that the cameras shoot the same scene. A specific method for calibrating the two linear cameras is presented in [12].

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. 6). 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.

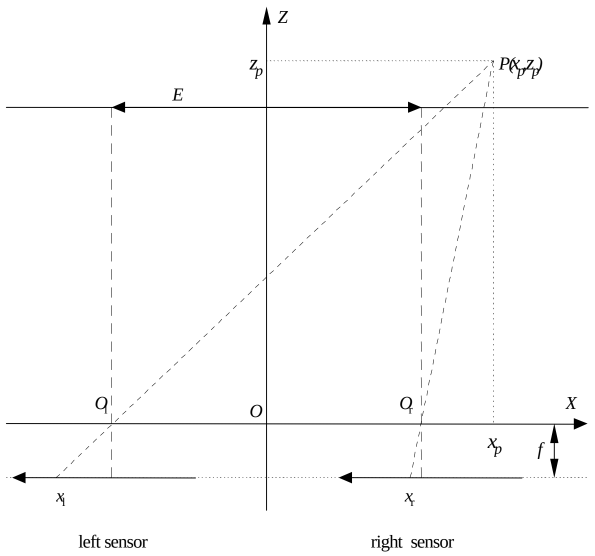


Fig.6 Pinhole lens model

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

$$z_p = \frac{E \cdot f}{d} \quad (3)$$

$$x_p = \frac{x_l \cdot z_p}{f} - \frac{E}{2} = \frac{x_r \cdot z_p}{f} + \frac{E}{2} \quad (4)$$

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

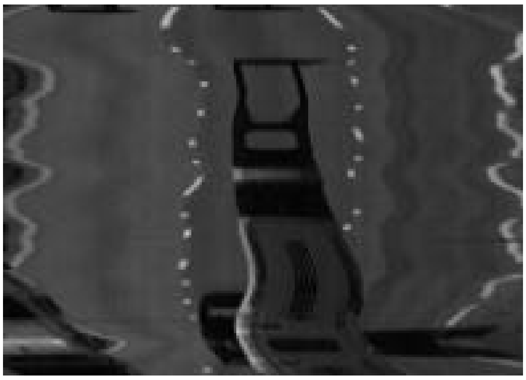
One of the sequences shot by the linear stereo set-up is shown in Fig. 7, where the linear images are represented as horizontal lines, time running from top to bottom.

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. The trajectories of the different vehicles during the sequence are shown in Fig. 8, where arrows indicate the movement of the three vehicles with respect to the prototype car.

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 prototype 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. The mechanical design of the stereoscope guaranties the stability of the calibration, even when the car is running on a rugged pavement.

Left sequence



Right sequence

Fig.7 Stereo sequence acquired by the set-up

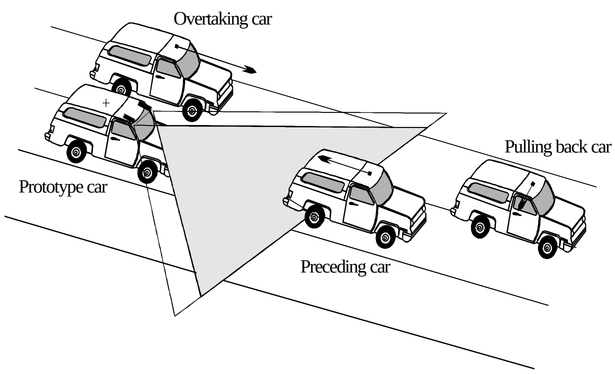


Fig.8 Trajectory of the different vehicles during the sequence

This stereo sequence has been processed using the proposed 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. 9, 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.

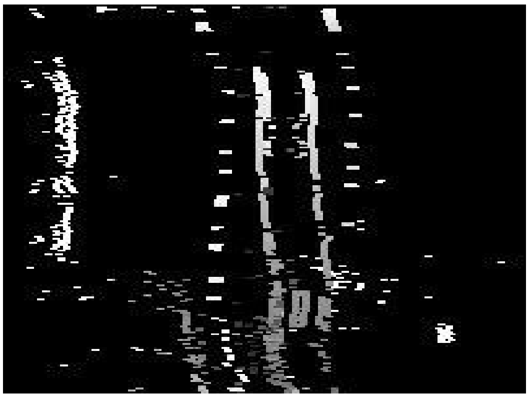


Fig.9 Reconstructed scene

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 vehicle, which pulls back in front of the preceding vehicle is identified as a white continuous line, at the bottom of the reconstructed image, corresponding to the left extremity of the vehicle. Its shadow can be see on the right 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 *80 ms*. The average processing rate is hence *2500* pairs of stereo linear images per second.

V. CONCLUSION

A voting based-method for real-time stereo correspondence is presented. The problem is first mapped onto a two dimensional matrix, called matching matrix, and then formulated in terms of finding pairs of true matches that satisfy local and global constraints. Local constraints are used to determine, from the matching matrix, the valid elements, which represent the compatible matches and participate to the voting process. Global constraints are used to define the rules, which govern the voting process. The performance of the proposed stereo matching method is evaluated for real-time obstacle detection in front of a moving car using linear stereo vision. The experiments show that the proposed method provides good results in terms of depth computation as well as processing rate.

VI. ACKNOWLEDGMENTS

We thank very much the French agency of innovation (Anvar) for its financial support. We also thank J.-G. Postaire, J.-C. Burie and J.-L. Bruyelle for their help and support.

VII. REFERENCES

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

[2] S. T. Barnard and W. B. Thompson, “Disparity analysis of images”, *IEEE Trans. on Patt. Anal. Mach. Intel.*, vol. 2, no. 4, 1980, pp. 333-340.

[3] U. R. Dhond and J. K. Aggarawal, “Structure from stereo - A review”, *IEEE Trans. on Syst. Man and Cybern.*, vol. 19, 1989, pp. 1489-1510.

[4] D. Marr and T. Poggio, “Cooperative computation of stereo disparity”, *Science*, vol. 194, 1976, pp. 283-287.

[5] S. Barnard and M. Fisher, “Computational stereo”, *ACM Comput. Surveys*, vol. 14, 1982, pp. 553-572.

[6] 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.

[7] 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*, vol. 11, 1993, pp. 261-268.

[8] R. M. Inigo and T. Tkacik, “Mobile robot operation in real-time with linear image array based vision”, *in Proceedings of the 1987 IEEE Int. Intelligent Control Symposium*, Philadelphia, USA, pp. 228-233.

[9] O. Colle, *Vision stéréoscopique à l’aide de deux caméras linéaires - Application à la robotique mobile*, Thèse de Doctorat, Institut National des Sciences Appliquées de Lyon, France.

[10] Y. Ruichek, J.-G. Postaire and J.-L. Bruyelle, “A neural approach for obstacle detection with a linear stereoscopic sensor”, *Mathematical an Computer Modelling*, vol. 27, no. 9-11, 1998, pp. 215-228.

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

[12] 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*, Thèse de Doctorat, Université des Sciences et Technologies de Lille, France, 1995.