Ground Plane Obstacle Detection using Projective Geometry

A.Branca, E.Stella, A.Distante
Istituto Elaborazione Segnali ed Immagini - CNR
Via Amendola 166/5, 70126 Bari Italy
e-mail: [branca,stella,distante]@iesi.ba.cnr.it

Abstract— **The problem addressed in this paper is obstacle detection in the context of mobile robot navigation using visual information. The goal is achieved by analyzing successive pairs of time varying images acquired with the TV camera mounted on the moving robot. Assuming the robot is moving on a flat ground, any obstacle is identified by any cluster of points not coplanar with the largest number of points lying on the ground plane.**

We identify a planar surface by a set of points characterized by the projective invariance of the cross ratio of any five points.

Our method recovers a planar surface by clustering high variance interest points characterized by invariance of cross ratio measurements in two differently projected images.

Once interest points are extracted from each image, the clustering process requires grouping of corresponding points preserving cross ratio measurements. We solve this twofold problem of finding correspondent points and grouping the coplanar ones through a global optimization approach based on classical nonlinear relaxation labeling technique imposing cross ratio invariance as constraint.

I. Introduction

An essential ability of an autonomous robot moving in an unknown environment on a ground plane is to detect the presence of obstacles in order to determine the free moving space.

A lot of existing systems accomplish this goal by using range information obtained from active sensors such as laser scanners, radars, and ultrasonics. Actually, vision-based solutions, although requiring much computational power, can provide much higher resolution than sonars.

The problem addressed in this paper is vision-based obstacle detection for an indoor mobile robot equipped with a single camera.

Vision-based approaches detect the obstacles or the free areas in the scene from a reconstructed model of the 3D space in front of the vehicle. Building of 3D structures can be performed through 3D Euclidean approaches, involving directly 3D primitives to recover 3D structures in 3D space, or 2D non-Euclidean approaches working on the projection in the image plane of the 3D structures. Due to the difficulties of computing reliable 3D primitives, the second class of approaches is more reliable. Classically, the problem of 3D reconstruction from 2D projections has been treated by computing the distance (depth) from the camera optical center to points in the world from their correspondences in a sequence of images using camera parameters. Actually, individual depth estimates are noisy and sparse, producing a few qualitative information useful for scene analysis, moreover this class of approaches require the vehicle motion or camera parameters to be known a priori, or the vehicle speed to remain constant.

Instead, the system we propose exploits the geometric properties of the scene and does not depend on the knowledge of camera parameters or vehicle motion.

Assuming that the ground floor is flat, the camera observes a planar surface in motion as the robot moves in the workspace. Floor obstacles are detected as deviations from planarity of the ground plane.

The work reported here is based on the analysis of pairs of successive images acquired by a TV camera mounted on a mobile vehicle. By imposing to the optical axis of the TV camera to intersect the ground plane while the vehicle is moving, the goal is to detect all obstacles lying on the ground plane as the surfaces which result not-coplanar with the largest observed planar region, i.e. the ground plane.

Our navigational problem does not require the precise knowledge of the 3D position of points in the scene, but only a qualitative and compact description of the environment.

Recent work focuses on the extraction of information from the images without using the calibration parameters, but only using results of projective geometry.

Our idea is to recognize floor-obstacles as the 3D structures projected in image plane regions easily detectable using some projective invariant constraints. In particular our aim is to use projective invariants to recover the planar structure of the ground plane, then any region not satisfying the imposed constraints is considered the image projection of floor-obstacles.

The simplest numerical property of a planar object that is unchanged under projection to an image is the cross ratio of five coplanar points. The geometric invariance of cross-ratio of five coplanar points has been just used in literature as constraint in tracking algorithms, planar region detection or object recognition but using probabilistic analysis [3], [13], [14], [4], [5], [7], [10]. In order to overcome the problems derived from the use of probabilistic decision rules ([9], [6]) we consider the projective invariance constraints into a global optimization process performing at same time matching and grouping of coplanar high interest feature points extracted from the analyzing images. By considering a large number of intersecting subsets of five points, obtained as combinations of available sparse features, the clustering problem is solved by searching for a solution in the space of all potential matches by imposing to satisfy the five-order constraint of cross ratio. Imposing the cross ratio invariance to be globally satisfied spares us to deciding about local measurements of cross-ratio, giving rise to a more robust approach.

Using graph theory, our clustering problem is equivalent to find the maximum clique (i.e. the largest subset of nodes mutually compatible) of an association graph whose nodes represent the potential feature matches and the 5-order links are weighted by cross ratio similarity. In literature maximum clique finding algorithms are essentially implemented as recursive graph search methods which are known to be NP-complete problems and of an exponential growth in computing time as the number of association graph nodes increases. On the other hand optimization techniques (hopfield neural networks [8], relaxation labeling [15]) have been resulted more reliable converging in polynomial times to an optimal solution. Actually, it has been shown ([2]) as an optimal solution to the general problem of *maximum clique finding* (i.e. finding the largest maximal clique) can be achieved through a simple version of the original relaxation labeling model of Rosenfeld, Hummel, Zucker.

We extend the classical relaxation labeling approach (based on binary compatibility matrices) to treat with compatibility matrices of order five (with coefficients determined through cross ratio similarity measurements) and apply it to our context of planarity detection by clustering of coplanar features. Since in our context we are not interested to an exact 3D reconstruction but only to a raw approximation of the surrounding environment with large planar surfaces, an optimization approach, as the non-linear relaxation labeling, is sufficient.

Summarizing, in our method, the high interest features, extracted using the Moravec's interest operator [11] from each frame of the sequence, are organized into appropriate relational graphs as nodes connected by 5-order links weighted by cross-ratio measurements (section 2). A planar region is represented by a set of totally connected nodes of the corresponding association graph and is determined through an optimization approach based on a non-linear relaxation labeling process (section 3). All feature points which results not coplanar with the largest recovered cluster (representing the ground plane) are considered as obstacles. The method performs at same time matching of features and clustering (section 4) based on coplanarity. In our experimental tests (section 5), we found our method to be very fast in converging to a correct solution, showing as despite common assumptions, higher order interactions help to speed-up the process.

II. Cross-Ratio Invariance

Clustering of coplanar features proposed in this paper is based on projective invariance of cross ratio of five arbitrary coplanar points. Our aim is to propose a global optimization approach to grouping coplanar features into clusters representing different planes presented in the scene without any a priori knowledge about the scene.
The cross-ratio is the simplest numerical property of an object that is unchanged under projection to an image. In the plane it is defined on four coplanar lines (u_1, u_2, u_3, u_4), incident at a single point (pencil of lines), in terms of the angles between them and is given by:

$$cr(\alpha_{13}, \alpha_{24}, \alpha_{23}, \alpha_{14}) = \frac{sin(\alpha_{13}) * sin(\alpha_{24})}{sin(\alpha_{23}) * sin(\alpha_{14})} \qquad (1)$$

where α_{ij} is the angle formed by the incident lines u_i and u_j.

Any five coplanar point set $X = (x_1, x_2, x_3, x_4, x_5)$ (not three of which collinear) will be characterized by the invariance of cross-ratio $cross(X)$ of the pencil of coplanar lines generated by joining a point of X with the other four.
Projective invariance of cross ratio imposes for each subset of five coplanar points X in the first image the corresponding points in the second image Y to have the same cross ratio.
Actually in our context we are dealing with features extracted from different unknown planes: given five arbitrary features, if the cross ratio is not preserved we cannot decide if the features are on different planes or they have been mismatched. In fact, in most previous works, using cross ratio invariance, assumptions about coplanarity or correctness of some matches are a priori made.
Our idea is to solve this twofold problem by imposing the cross ratio invariance constraint to be satisfied on a lot of combinations of available features. The goal is to recovering large subsets of features all mutually compatible (i.e. correctly matched and coplanar) by imposing cross ratio invariance to be globally satisfied. Our problem is then reduced to a global constraint satisfaction problem to be solved through an optimization approach selecting sets of matches mutually compatible, i.e. pairs of features correctly matched and coplanar.

III. High-Order Relaxation Labeling Optimization

Relaxation labeling processes provide an efficient optimization tool to solve difficult constrained satisfaction problems, making use of contextual information to solve local ambiguities and achieve global consistency [2].

A general labeling problem involves a set of objects to be labeled through a set of labels, using local measurements of objects and contextual information recovered from knowledge of relations between objects. This is quantitatively expressed by a compatibility coefficient matrix.

In our context, the two frames of the sequence to be matched are represented by two relational graphs, characterized by nodes $(\{n1_i\}, \{n2_j\})$ representing the feature points extracted $(\{p_i\},\{q_j\})$ using the Moravec's interest operator and five-order links $(\{L1_{i_1 i_2 i_3 i_4 i_5}\}, \{L2_{j_1 j_2 j_3 j_4 j_5}\})$ weighted by the cross-ratio $(CR1_{i_1 i_2 i_3 i_4 i_5} = cross(p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}, p_{i_5})$, $CR2_{j_1 j_2 j_3 j_4 j_5} = cross(q_{j_1}, q_{j_2}, q_{j_3}, q_{j_4}, q_{j_5}))$ evaluated on the features associated to the five connecting nodes.

Optimal matches should be recovered through a search on the association graph $G = \{\{n_h\}, \{l_{hklmn}\}\}$ consisting of nodes $\{n_h\}$ representing candidate matches and five-order links $\{l_{hklmn}\}$ weighted by cross ratio similarity (2).

$$C_{hklmn} = e^{-(|CR1_{h_i k_i l_i m_i n_i} - CR2_{h_j k_j l_j m_j n_j}|^2)} \qquad (2)$$

The goal is to select all matches mutually compatible according to compatibility matrix C.

Theoretically, the association graph G should consists of $N \times M$ nodes $\{n_h\}$, representing candidate matches $\{(p_{h_i}, q_{h_j})\}$ and $\binom{N \times M}{5}$ links weighted by compatibility matrix values $\{C_{hklmn}\}$. Actually, in oder to reduce the number of links involved in the process, we generate a node n_h only if the corresponding features p_{h_i} and q_{h_j} have an high radiometric similarity. The correct set of matches is found by an iterative nonlinear relaxation labeling process. A relaxation labeling process takes as input an initial labeling (matching) assignment and iteratively updates it considering the compatibility model. In our problem, to each node n_h is assigned the initial labeling Λ_h representing the degree of confidence of the hypothesis p_{h_i} *is matched with* q_{h_j}. The relaxation algorithm updates the labeling $\{\Lambda_h\}$ in accordance with the compatibility model, by using a *support function* γ (3) quantifying the degree of agreement of the h-th match with the context.

$$\gamma_h = \sum_{klmn} C_{hklmn} \Lambda_k \Lambda_l \Lambda_m \Lambda_n \qquad (3)$$

The rule of adjustment of the labeling $\{\Lambda_h\}$ should increases Λ_h when γ_h is high and decreases it when γ_h is low. This leads to the following updating rule:

$$\Lambda_h = \Lambda_h \gamma_h / \sum_k \Lambda_k \gamma_k \qquad (4)$$

Iteratively all labeling of $\{n_k\}$ nodes are updated until a stable state is reached. In the stable state all $\{n_h\}$ nodes with non-null labeling Λ_h will represent optimal matches between coplanar features. In [2] it has been shown that this algorithm posses a Liapunov function. This amounts to stating that each relaxation labeling iteration actually increases the labeling consistency, and the algorithm eventually approaches the nearest consistent solution. The solution is the largest set of nodes mutually compatible according to compatibility matrix $\{C_{hklmn}\}$, i.e. the maximum clique identifying the largest planar face. In order to recover all planar faces in which the available features can be organized, we should repeat the relaxation process iteratively, as described in the next section.

IV. Clustering

In our context, each plane is characterized by a set of assignment nodes totally connected as specified by the imposed projective invariant constraint of cross-ratio. The goal is to determine all clusters of nodes mutually compatible. In graph theory a set of nodes mutually compatible is known as clique. A clique is called maximal if no strict superset of it is a clique. A maximum clique is a clique having largest cardinality.
In [15] it is shown how the relaxation model of Rosenfeld, Hummel and Zucker [1] above described is capable of approximately solving the maximum clique problem.

Our problem is solved by recovering all maximal cliques identifying all different planes. The relaxation labeling approach gives us an approximation of the maximum clique: i.e. the largest set of features correctly matched and coplanar, this amounts to estimate the planar surface for which a greatest number of features have been extracted. Actually our goal is to recovering as many subsets of mutually compatible nodes as the large planar surfaces are observed in the scene. This can be accomplished by applying iteratively the relaxation algorithm until all features have been clustered. At each iteration, the current largest subset of feature matches mutually compatible under the cross ratio invariant constraint is identified as the maximum clique and it is pruned from the association graph. When all maximal cliques have been identified the remaining graph nodes will represent mismatched features or sparse isolated features.

We point out that the method, while finding the maximal cliques, recovers at same time all correct matches.

Finally, in order to reduce the computational complexity (of order $\binom{N}{5}$, if N is the number of involved match nodes), for each cluster to be determined, instead to apply the relaxation process for maximum clique finding to the whole set M of the N involved matches, we consider m smallest subsets $\{M_1, \ldots M_m\}$ - each of n matches randomly selected from M ($n \ll N$)- so that $M_i \cap M_j \neq \emptyset$ and $M_i \subset M$ (for $i, j = 1, ..m$) and $m\binom{n}{5} \ll \binom{N}{5}$. The maximum clique of the whole set M is generated by merging all maximum cliques estimated from all smallest subsets $\{M_1, \ldots M_m\}$.

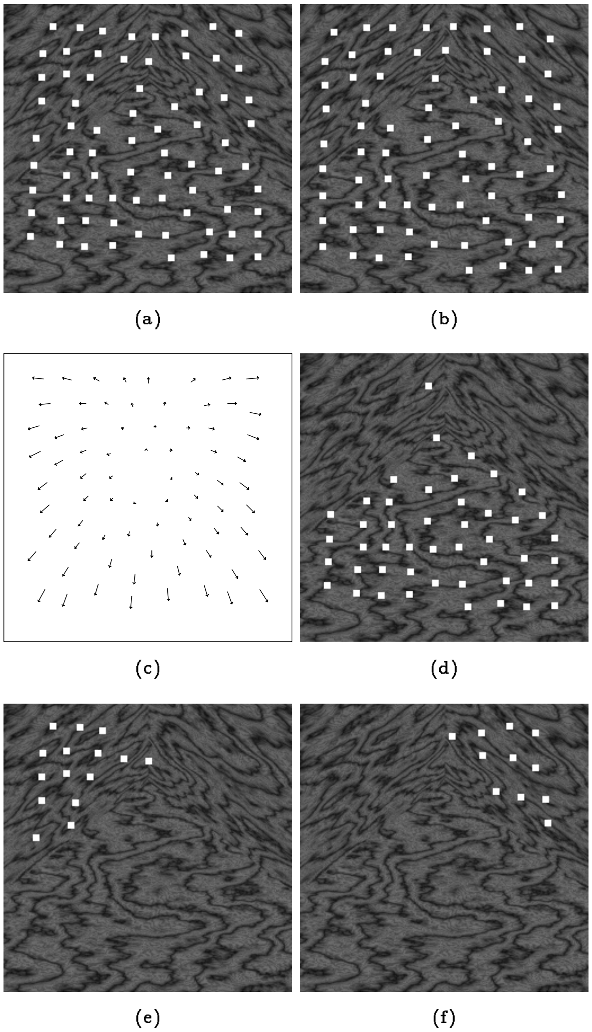


Fig. 1. A synthetic time varying image sequence. The simulated scene is constituted of three orthogonal planes (one horizontal (α), and two vertical (β,θ)). The simulated camera motion is a forward translation along a direction parallel to α while the camera axis is rotated of an angle of $45deg$ with respect to the moving direction. (a)(b) are the two analyzed consecutive images with superimposed the extracted features. (c) is the map of feature correspondences as estimated by the radiometric similarity. (d)(e)(f) are the recovered planes: α, β and θ, respectively

The so clustered mutually compatible features are pruned from M and the process is repeated on the remaining.

V. Experimental Results

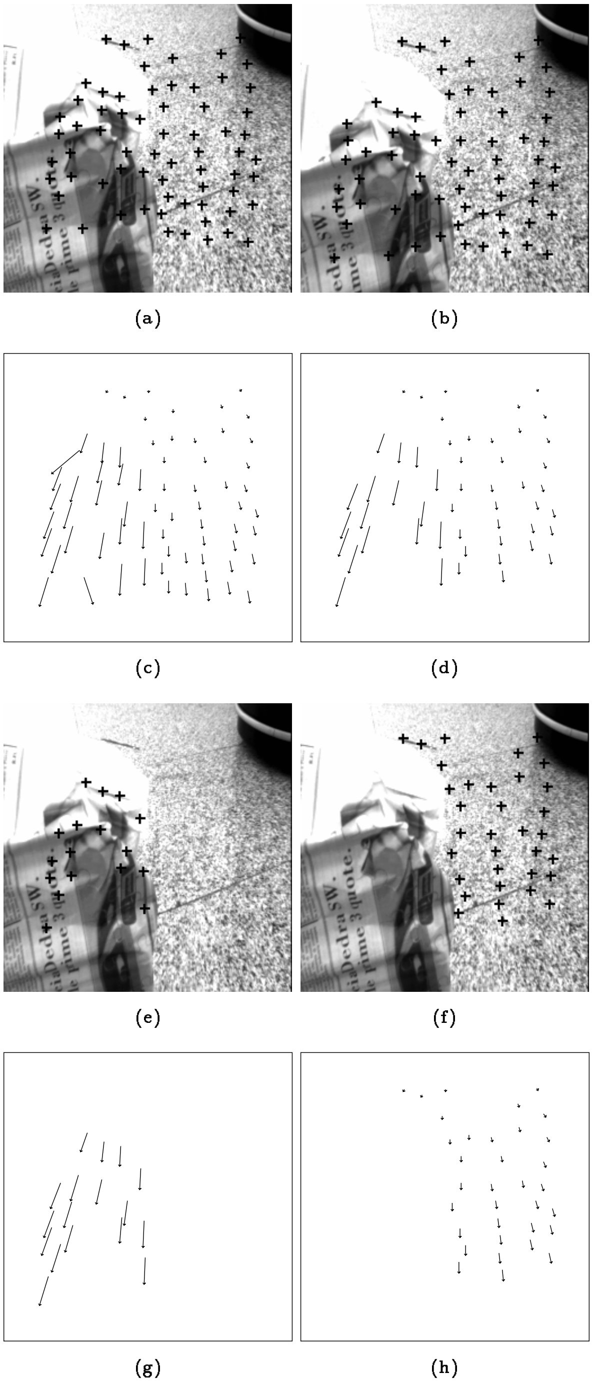
The first experiment reported in figure (1) is about the results obtained while testing our method in different complex contexts generated synthetically. The goal was to measure the ability to cluster coplanar features of planes apparently indistinguishable because of their similar texture and approximatively similar distance from the optical center of the TV camera. In fact, our aim was to test our method in contexts where most of common floor obstacle detection approaches (often based on floor recognition or computation of depth differences) fail.

In particular most of previous approaches detecting obstacles by analyzing maps of feature correspondences are based on the estimation of the Time To Collision (TTC) from the estimated 2D or 3D velocities. Such methods, assuming the linear approximation of the motion field, fail when the obstacle is tilted, or the distance from the camera optical center is large with respect the focal length, or the obstacle is located in a small portion of the acquired images.

In the experiments reported in fig.(2-3), we can observe as our method is able to detect the obstacles, even in these difficult contexts. Figures (2-3) report some results obtained on real time-varying image sequences acquired in our laboratory by a TV camera (6mm focal length) mounted on a *Nomadic Scout* mobile vehicle. The vehicle was constrained to move forward by following rectilinear paths along which different obstacles were located on the floor. The TV camera was inclined so the optical axis intersects the floor at a distance from the vehicle of approximatively 50cm. The two experiments here reported differ each other by the distance and orientation of the obstacle with respect the optical axis. In fig.(2) the obstacle is nearest the camera and the normal to its visible surface is parallel to the direction of translation of the vehicle. In this context the features on the obstacle appear most near to the vehicle than the features on the floor, as evidenced by their vector modules. On the contrary, in fig.(3) the obstacle is farthest and its surface is tilted, then no significative differences can be observed among the velocity vectors of the features of the different planar surfaces.

Though in the first case (fig.(2)) one can think to solve the problem by means of simple calculations based on 2D velocity measures, in the second case (fig.(3)) such approach will fail.

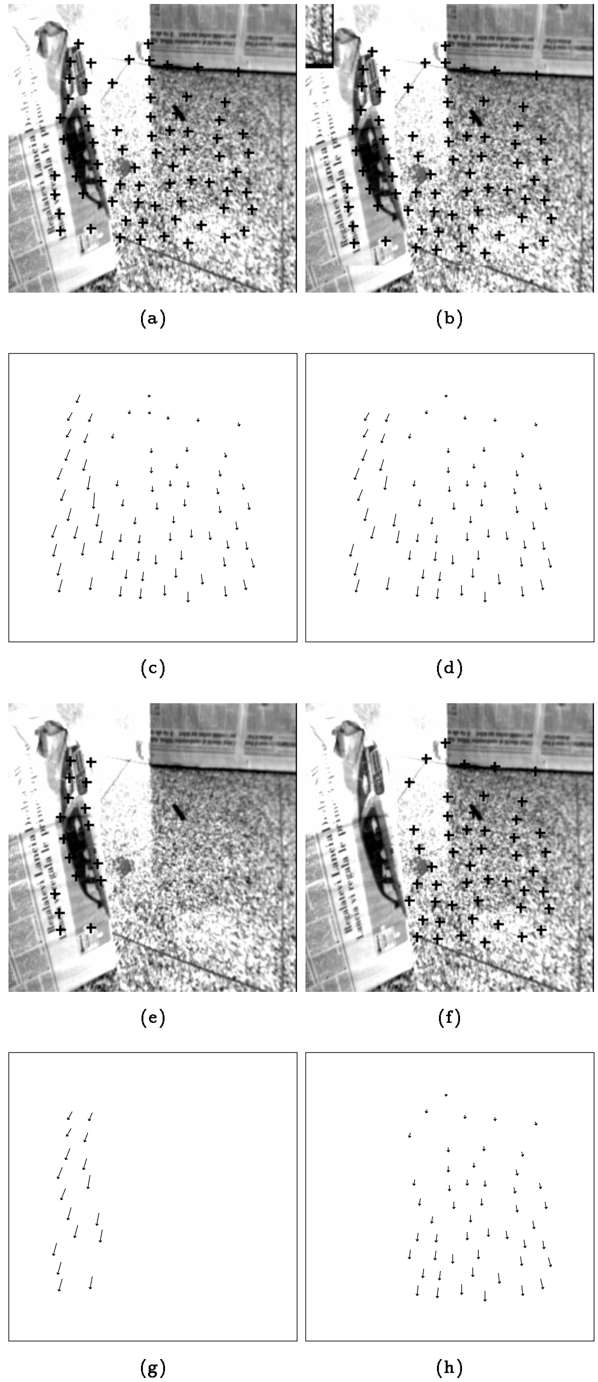
Moreover, it is well known as the 3D velocities estimated on small regions far from the singular points (i.e. the points where the motion field vanishes, then in this context the point where the optical axis intersect the image plane) cannot provide robust information to perform a correct classification of the 3D motion of the vehicle. Consequently considering 3D velocity to estimate (even qualitatively) the TTC of the vehicle with respect to different portions of the observed scene is not an efficient approach. This can be verified by observing the values of the 3D translational motion parameters (T_X,T_Y,T_Z) (fig.(2.i)-(3.i)) of the vehicle estimated (using the approach of [16]) along the three orthogonal axis (X, Y, Z) by analyzing the feature correspondence maps relative to the whole image (before the optimization and after) and the different recovered planar surfaces. The motion of the vehicle is a forward translation, however, because the camera is inclined so its optical axis intersects the the floor, the 3D motion should be described by only T_Y and T_Z.



	Tx	Ty	Tz
(c)	-0.0142	0.0816	-0.0503
(d)	-0.0147	0.0810	-0.0577
(g)	-0.0243	0.1528	-0.0399
(h)	-0.0073	0.0442	-0.0351

(i)

Fig. 2. (a)(b) two analyzed consecutive real images with superimposed the extracted features. (c) the map of feature correspondences as estimated by the radiometric similarity. (d) the final map of optimal feature correspondences. (e)(f) the two distinct clusters of features points representing respectively the detected obstacle and the floor. (g)(h) the two recovered clusters of feature correspondences. (i) the estimated 3D velocities.



	Tx	Ty	Tz
(c)	-0.0048	0.0485	-0.0254
(d)	-0.0048	0.0480	-0.0245
(g)	-0.0140	0.0649	-0.0146
(h)	-0.0045	0.0413	-0.0269

(i)

Fig. 3. (a)(b) two analyzed consecutive real images with superimposed the extracted features. (c) the map of feature correspondences as estimated by the radiometric similarity. (d) the final map of optimal feature correspondences. (e)(f) the two distinct clusters of features points representing respectively the detected obstacle and the floor. (g)(h) the two recovered clusters of feature correspondences. (i) the estimated 3D velocities.

VI. Conclusions

In this paper, an obstacle detection approach using only some correspondences estimated among features extracted from two successive time varying images acquired while an autonomous vehicle is moving on a flat ground has been developed. If the camera optical axis intersect the ground plane all obstacles are easily detectable as the points non coplanar with the largest set of coplanar feature points extracted on the ground plane. The method don't require a priori knowledge about vehicle motion and structure environment, or coordinate transformations. Through an optimization process the free space on the ground plane is detected as the largest cluster of points which result coplanar under the constraint of projective invariance of cross ratio of five coplanar points.

As verified through the above experimental tests, with respect to classical obstacle detection approaches based on TTC, our method provides most useful and robust informations about the the location of the obstacle, even at a largest distance where significant depth differences are not appreciated.

The required computational time still depends on the number of extracted features and the cardinality of subsets considered to estimate a maximum clique. In our experiments, we find the minimum cardinality of such small subsets $\{M_1, ...M_m\}$ is 15, and a maximum of 30 iterations are necessary to relax (i.e. to reach a stable state the relaxation algorithm) requiring 0.2-0.3 seconds. Since in real contexts the number of large surfaces to be recovered are no more than two or three, the required computational time to perform the whole process is of the order of 3-4 seconds at maximum. We should observe that this time can be additionally optimized since a lot of process are independent and can be performed in parallel.

Moreover, if the goal is to obtain only the free space on the ground plane, the process can be stopped when the first largest cluster (i.e. the first maximum clique) is detected, so allowing to performs all calculations in real time.

References

[1] A.Rosenfeld, R.A.Hummel, and S.W.Zucker "Scene labeling by relaxation operations" *IEEE Trans. Syst. Man. Cyber.* 6(6):420-433, 1976.

[2] M.Pelillo "The Dynamics of Nonlinear Relaxation Labeling Processes" to appear in *Journal of Mathematical Imaging and Vision* September 1996.

[3] H.Chabbi, M.O.Berger "Using Projective Geometry to Recover Planar Surfaces in Stereovision" *Pattern Recognition* Vol.29, No.4, pp.533-548, 1996.

[4] S.Carlsson "Projectively Invariant Decomposition and recognition of Planar Shapes" *International Journal of Computer Vision* Vol 17 No 2 pp 193-209 1996.

[5] C.A.Rothwell,A.Zisserman,D.A.Forsyth,J.L.Mundy "Planar Object Recognition using Projective Shape Representation" *International Journal of Computer Vision* Vol 16 pp 57-99 1995.

[6] S.J.Maybank "Probabilistic Analysis of the Application of Cross Ratio to Model Based Vision" *International Journal of Computer Vision* Vol 16 pp 5-33 (1995)

[7] D.Oberkampf, D.F.DeMenthon, L.S.Davis "Iterative Pose Estimation using Coplanar Feature Points" *Computer Vision and Image Understanding* Vol 63 No 3 pp 495-511 1996.

[8] A. Jagota "Approximating Maximum Clique with a Hopfield Neural Network " *IEEE Transaction on Neural Networks* Vol.6, No.3, pp.724-735, May 1995.

[9] S.J.Maybank "Probabilistic Analysis of the Application of Cross Ratio to Model Based Vision: Misclassification" *International Journal of Computer Vision* Vol 14 pp 199-210 (1995).

[10] K.Kanatani "Computational Cross Ratio for Computer Vision" *CVGIP: Image Understanding* Vol 60 No 3 pp 371-381 1994.

[11] H.P.Moravec *The Stanford Cart and the CMU Rover*,Proc. IEEE,1983.

[12] R.A Hummel and S.W. Zucker "On the foundations of relaxation labelling processes" *IEEE Trans. Patter. Anal. Mach. Intell.* Vo. 5, N. 3, 1983, 267-287.

[13] D.Sinclair, A.Blake "Qualitative Planar Region Detection" *International Journal of Computer Vision* Vol 18 No 1 pp 77-91 (1996).

[14] P.Gurdjos, P.Dalle, S.Castan "Tracking 3D Coplanar Points in the Invariant Perspective Coordinates Plane" *Proc. of ICPR'96* Vienna 1996.

[15] M.Pelillo "Relaxation Labeling Networks for the Maximum Clique Problem" *Journal of Artificial Neural Networks* 2(4), 313-328 (1995).

[16] A.Branca, G.Convertino, E.Stella, A.Distante "A Neural Network for Ego-motion Estimation from Optical Flow" Proc. BMVC95.