Towards Cooperative Visual-Based Localization, Mapping, and Servoing

Antonio Danesi Daniele Fontanelli Antonio Bicchi

Abstract— **In this paper we consider the problem of navigating an autonomous robot using primarily vision for localizing the robot, building a map of the environment, and navigating through viapoints to goals. This can be regarded as an attempt to bridge the gap between existing techniques for image–based localization and mapping, and methods for visual servoing of mobile robots, i.e. between robotic perception and action.**

A visual servo scheme is used that can steer the wheeled vehicle among images. Goal images do not necessarily correspond to images physically taken from the desired vehicle posture, as servoing to reconstructed virtual images is possible. A topological image map is constructed to support this, based on images grabbed by on-board cameras, along with a global feature-based metric map, using extended Kalman filter techniques. The method also enables a team of multiple vehicles to merge their information, and to coordinate navigation using each other's images.

Realistic assumptions on limited communication bandwidth between agents and available memory storage are taken into account considering informative, memory-safe maps. Simulations and preliminary experimental results on a laboratory setup are reported.

I. Introduction

One of the main obstacles that still hinder penetration of mobile robots into wide consumer markets is the unavailability of powerful, versatile and cheap sensing. Vision technology is potentially a clear winner as far as the ratio of information provided versus cost is considered: cameras of acceptable accuracy are currently sold at a e which is one to two orders of magnitude less than laser scanners. As a consequence, much attention is being devoted to solving the non–trivial problems implied by using visual information for building maps, localizing and navigating through the environment.

This paper deals with the problem of using off-the-shelf cameras fixed on inexpensive mobile platforms, to enable navigation and control to given goal configurations in space based on visual maps of the unknown environment, which are contextually built in the process. To this purpose, some powerful tools have been provided recently in the literature on localization and mapping for autonomous vehicles mainly by CS and AI techniques, and separately by research work on visual servoing of robots, coming mostly from an automatic control background. Our effort is mainly focused on the integration between advanced techniques of sensing and understanding the ambient (perception), and the necessity of making and implementing decisions (action) based on real-time sensorial feedback fromthe environment.

The problem of simultaneous localization and map building, or SLAM ([6], [24], [37]), and the control of a vehicle are clearly both intrinsically sensor-related. Nevertheless, control aspects are often disregarded in the classical approach to SLAM, whereby the map is most often built off–line and with no explicit reference to the usage of data that will be done to close the loop on the controller actions. On the other hand, localization (i.e., full state information) of the vehicle is often taken for granted in the rather extensive autonomous vehicle control literature ([25], [8], [11], [7]). In system-theoretic terms, such decoupling of the estimation problem from the feedback design problem seems to appeal to a generalization of the *separation principle* of linear stochastic control, which unfortunately is a mere leap of faith in the context of robotic systems with highly nonlinear dynamics and sensors (see e.g. [5] for a discussion of SLAM from a system-theoretic viewpoint).

A. Previous work

Visual servoing of vehicles can be regarded as an attractive shortcut for the estimation/control problem, implementing feedback directly on output measurements, i.e. grabbed images. Previous work has considered the case of a camera that can move independently from the vehicle ([17]), or is carried by an articulated arm mounted on the robot ([41]). The more economically viable solution of fixed onboard cameras has been considered e.g. by [16], [10]. Furthermore, [30] considered the practically most relevant problem of keeping features to be tracked within sight of a limited–aperture camera while the vehicle maneuvers, and proposes a hybrid controller that solves the stabilization problem. A modified version of the controller in [30], whereby the technique is generalized to require no a-priori information on the 3D environment, is also adopted as a component of the architecture proposed in this paper.

To achieve the goal of an autonomous system capable of navigating in a complex environment using only vision (along with odometry when available), the capability of servoing to a given prespecified image is clearly not enough. The system should also build and update maps for the environment, in terms that are both informative enough for control tasks (to allow e.g. a vehicle to accurately reach arbitrary positions in the environment), and aware of resource limitations (such as memory space, or communication bandwidth for multiple robot systems). In this sense, our work is indebted with the SLAM literature, particularly for vision-based localization and mapping ([35], [33], [18] [28], [23]). Cooperative localization and map build-

Support from EC contract IST2001–37170 "RECSYS" and from MIUR under grants PRIN 095297_002-2002, FIRB RBAU01RY47.

Authors are with the Interdept. Research Center "Enrico Piaggio", University of Pisa, via Diotisalvi, 2, 56100 Pisa, Italy. Phone: +39050553639. Fax:+39050550650. E–mail: danesi,fontanelli,bicchi@piaggio.ccii.unipi.it

ing by multiple agents may allow minimizing the time for building complex maps and incrementing the accuracy of the produced maps with augmented fault tolerance ([12], [22], [33], [34], [39]).

A relatively novel approach in robot control uses *Augmented Reality*, based on virtual representation of the environment taken from virtual cameras. In [31] an automatic generator of camera motion using motion planning techniques is presented to help users navigate in virtual environments specifying only the starting and final position of the virtual camera. In [29], a virtual environment containing objects' models and a robotic arm with eye-in-hand camera is used to train the visual based controller. Virtual visual servoing [27] combines augmented reality techniques with the servoing architecture for pose computation, by modifying the parameters of a virtual camera (position, orientation, and if necessary intrinsic parameters) using the visual servoing paradigm. Other useful applications are to non-linear camera calibration ([26]) or to robust pose estimation with respect to marker partial occlusions by virtual objects ([9]).

At a higher level, reasoning about planning necessitates the abstraction (or "anchoring") of metric data to logic entities or "symbols" (such as rooms, corridors, doors). Topological approaches have been widely investigated in the computer science community, often in connection with studies on cognitive maps in human learning processes ([42], [32], [15], [19]). The seminal paper of Kuipers on Spatial Semantic Hierarchy (SSH) ([20]) defines a useful architecture for robot navigation using different semantic layers, comprehensive of a cognitive map builder and a fuzzy navigation controller. In [21], [4], [14] the SSH approach has been adopted and adapted to different problems. In the literature of mobile robots control, Dudek et al.. have introduced robot graph navigation [13] and a global map based on a set of reliable and accurate local maps organized in a topological structure in [36]. Thrun et al. have studied this particular problem in [38], integrating grid-based maps and topological maps generated partitioning the former into coherent regions. In [40], the map process is split in two phases: a topological mapping phase solves the global position alignment problem while a metric mapping phase produces a fine-grained metric map of the environment.

B. The V-SLAMS architecture

The system we are developing to address the problem of visual-based simultaneous localization and mapping for servoing (V-SLAMS) is comprised of several interconnected components. The final goal of our project is to have multiple mobile agents, equipped with cameras and possibly basic odometry, cooperatively build a visual map of the environment. The map should be such as to allow any single vehicle to localize itself, and navigate through the map to reach an arbitrary position, which may not have been reached in advance. The mapping and servoing phases should not necessarily be thought as consecutive in time.

In our proposed architecture, sensorial data organization

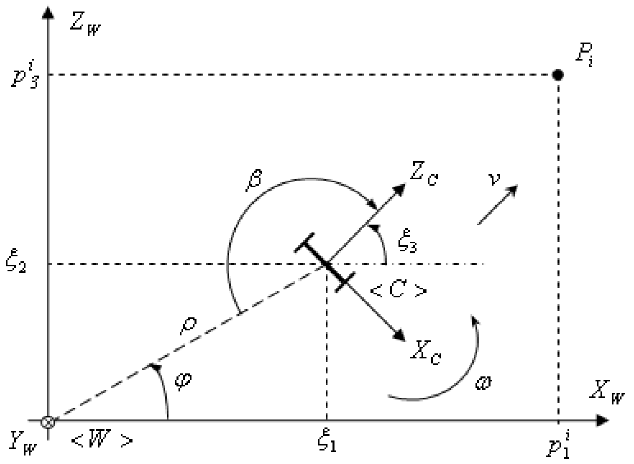

Fig. 1. Fixed frame $< W >$, camera frame $< C >$, and relative coordinates (ξ_1, ξ_2, ξ_3) and (ρ, ϕ, β)

starts with stochastic estimation of the 3D coordinates of image features, and possibly of uncertain parameters in the camera and environment models, via Extended Kalman Filter (EKF) techniques. Estimated values are merged into a general 3D feature map, useful for robot navigation and localization. The flexibility of the feature map allows a rather simple improvement of the extracted information using standard computer vision algorithms, like texture extraction and homography plane detection (see [12] for preliminary results on this topic using a cooperative architecture). A topological image-based map is also maintained in parallel, which is induced by the used architecture, to effectively connect the feature-based map and the topology of the surrounding ambient. Once vehicles are localized with respect to any point of the map, cooperation is enabled by sharing the global feature-based map and the image map, allowing robots to regroup in any position with respect the environment or relative vehicles position. Virtual visual servoing based on reprojected desired images and features can then be used in our scheme to accurately place the vehicle even in unexplored portions of the environment. Particular attention is to be devoted in the implementation of the V-SLAMS architecture to produce maps whose requirements in terms of memory allocation and communication bandwidth for sharing are limited.

In this paper, we describe our preliminary work towards the solution of the cooperative V-SLAMS problem. We use one or more vehicles with fixed cameras on-board, and build upon a combination of previous results on visual servoing techniques. Simulations and preliminary experimental results on a laboratory vehicle are reported, showing the practicality of the proposed approach.

II. Building the maps

We consider vehicles equipped with a fixed calibrated monocular camera, whose axis is aligned with the forward motion direction of the vehicle (which will be assumed to

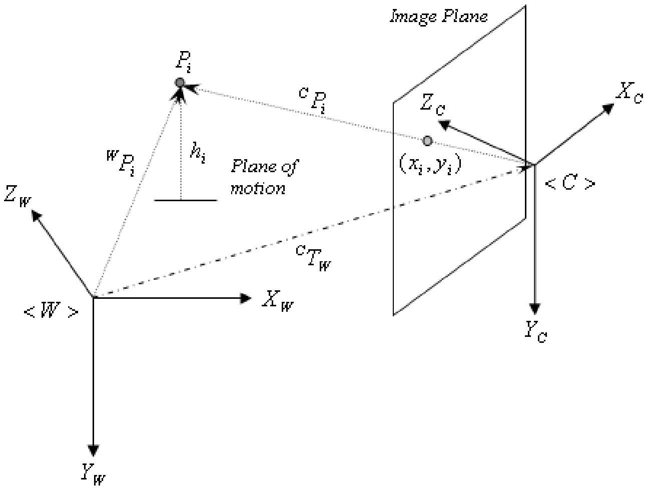

Fig. 2. Fixed frame $< W >$, camera frame $< C >$ and relative feature coordinates.

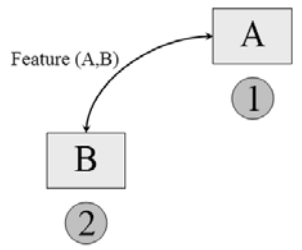

Fig. 3. Topological image map. Grabbed images are indicated with a capital letter, say A, B. The corresponding numbers are the 3D vehicle position with respect to the relative image.

behave like a kinematic unicycle). A moving camera frame $< C >$ is rigidly placed on the mobile robot chassis with the origin in the camera pinhole, with the Z_c axis directed along the camera optical axis and with the Y_c axis perpendicular to the plane of motion and passing through the middle point of the unicycle axle (see fig. 1). From the current image, $n \geq 3$ characteristic points (features) are selected with corresponding coordinates in the camera frame ${}^C P_i = {}^C[p_1^i, p_2^i, p_3^i]^T$. The vehicle motion is assumed to be constrained on the ${}^C X \times {}^C Z$ plane. This hypothesis is justified when the robot moves at a fixed level, e.g. on the floor of an office or factory space, and implies that the coordinates ${}^c p_2^i = h_i$ of each feature are constant and represent the height of the feature on the plane of motion (see fig. 2).

The position of each feature in the image plane is described by the well known perspective projection mapping $\Upsilon : \mathbb{R}^3 \to \mathbb{R}^2$

$$\Upsilon \ : \ {}^C P_i \ \rightarrow \quad \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \alpha_x \frac{{}^C p_1^i}{{}^C p_3^i} \\ \alpha_y \frac{h_i}{{}^C p_3^i} \end{bmatrix}. \qquad (1)$$

where (x_i, y_i) are the feature coordinates in the image plane. α_x and α_y are camera calibration parameters that represents the focal length and the pixel dimension scale factor on the image.

Consider now a fixed frame $< W >$ whose origin is coincident with the origin of $< C >$ when the robot is in the initial configuration, and with $X_w = Z_c$ and $Y_w = Y_c$. Let ${}^W \xi = {}^W[\xi_1, \xi_2, \xi_3]^T \in \mathbb{R}^2 \times S$ denote the robot posture. More precisely, (ξ_1, ξ_2) are the cartesian coordinates of the middle point of the unicycle axle, and ξ_3 is the orientation of the unicycle between the Z_c axis and the X_w axis, as represented in figure 1. From the initial unknown position of the vehicle (i.e. ${}^W \xi = {}^W[0, 0, 0]^T$) an image of a portion of the scene in view is grabbed and stored as the first image of the topological image-based map (see fig. 3, image A taken from starting position 1). From the image in view, a subset of features are selected following some standard cost index, such as minimum distance between two features, luminosity, or tracking simplicity. Let ${}^W P_i = {}^W[p_1^i, p_2^i, p_3^i]^T$ be the i–th feature coordinates with respect to $< W >$ – note that all the features are motionless in $< W >$. The same feature coordinates in the initial camera frame $< C >$ are: ${}^C P_i = {}^C[p_1^i, p_2^i, p_3^i]^T = {}^C[-{}^W p_3^i, {}^W p_2^i, {}^W p_1^i]^T$. At this point, a simple image-based control law is implemented allowing the robot to track the feature while it moves in any direction. An Extended Kalman Filter (EKF) is implemented on camera measurements to estimates the relative spatial position of the feature in camera frame $< C >$. It is worthwhile to note that localization of the robot can be disregarded during this initial step, being it sufficient that the robot keeps viewing and tracking the features, thus uncorrelating the estimation process for each feature and eliminating outlier mismatch possibilities. The estimated EKF state is $S^f = [S_1^f, S_2^f, S_3^f, \ldots, S_{3n-1}^f, S_{3n}^f]^T = [{}^C p_1^1, {}^C p_2^1, {}^C p_3^1, \ldots, {}^C p_2^n, {}^C p_3^n]^T$, i.e. the n features coordinates to estimate in the $< C >$ camera frame. The kinematic model is computed for state prediction assuming the camera fixed as specified in figure 1, obtaining

$$\begin{bmatrix} \hat{S}_1^f(k+1) \\ \hat{S}_2^f(k+1) \\ \hat{S}_3^f(k+1) \\ \vdots \\ \hat{S}_{3n-1}^f(k+1) \\ \hat{S}_{3n}^f(k+1) \end{bmatrix} = \begin{bmatrix} \hat{S}_1^f(k) + \hat{S}_3^f(k) u_2(k) + w_1(k) \\ \hat{S}_2^f(k) + w_2(k) \\ \hat{S}_3^f(k) - u_1(k) - \hat{S}_1^f(k) u_2(k) + w_3(k) \\ \vdots \\ \hat{S}_{3n-1}^f(k) + w_{3n-1}(k) \\ \hat{S}_{3n}^f(k) - u_1(k) - \hat{S}_{3n-1}^f(k) u_2(k) + w_{3n}(k) \end{bmatrix},$$

where $U(k) = [u_1(k), u_2(k)]^T$ are the encoder measurements for forward and angular velocity, obtained from $u_1(k) = R\frac{\omega_r(k)+\omega_l(k)}{2}$ and $u_2(k) = R\frac{\omega_r(k)-\omega_l(k)}{L}$ respectively. ω_r and ω_l are the rotational encoder for the right and left wheel, R is the wheel radius and L is the length of the wheel axle. $w = (w_1, w_2, w_3, \ldots, w_{3n-1}, w_{3n})^T$ is the additive zero mean gaussian noise comprised of model dynamical errors and noisy odometry measurements. Perspective projection inversion acts as the filter measurement correction, again added with zero mean gaussian noise representing image coordinates extraction inaccuracy. Covariance error matrices are weighted taking into account odometry lack of accuracy (typically due to wheels slipping and skidding).

Once feature position estimations have converged to the "real" desired value, that is features' positions known with

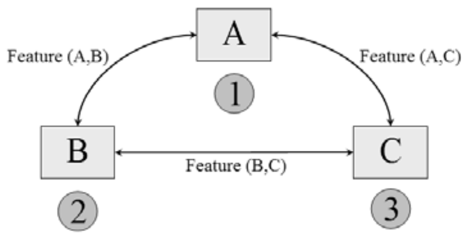

Fig. 4. Topological image map. Graph nodes represent grabbed images. Arcs connecting two image nodes are marked with the corresponding estimated feature, useful for visually servoed navigation.

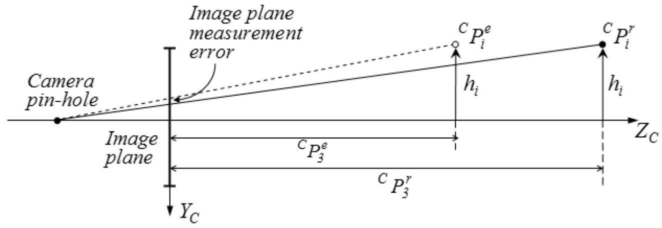

Fig. 5. Perspective projection for Z_c optical axis direction. The point ${P_i}^r$ is the 3D feature real position. Erroneous image coordinates measurements will corrupt 3D feature estimation, wrongly positioning the feature in ${P_i}^e$. The error is magnified if the corresponding y_i image coordinate approaches the zero (3D feature near the horizon).

a low level of uncertainty, a grabbed image is inserted in the topological image map with the relative vehicle position, and estimated feature coordinates are inserted in the feature-based map. In figure 3, the image B grabbed from position 2 is joined with A with an arc representing the subset of estimated feature. The wheeled mobile robot localization of position 2 in $< W >$ fixed frame is achievable knowing the initial image A, more precisely knowing $^{C}P_i$ with $i = 1..n$ in B and the feature correspondences in A. In this way the vehicle is able to be visually servoed between image A and B.

From image A or B (i.e. from position 1 or 2) the algorithm is repeated, adding new image to the topological image map and new estimated feature to feature-based map (see fig. 4). Topological map images are hence regarded as graph nodes, connected each other by feature arcs (see fig. 4), representing the necessary visual servoing controller features. It is worthwhile to note that a single connection between two nodes is used to travel back and forth the node images (desired image and final image are interchangeable), so that graph navigation is dead-lock free.

The described structure is ready-to-use with standard graph visiting algorithm, permitting path selection in the image graph to steer the vehicle in the environment, possibly with a cost function, such as minimum time, minimum distance or minimum control effort, implemented by simply associating weights to arcs. The presented representation implies two different representations of the vehicle state: a point in the continuous space of planar configurations $SE(3)$, and a discrete state in the topological graph.

III. Visual Servoing

The baseline of the visual scheme adopted within this work is the hybrid Position-Based Visual Servoing scheme presented in [30]. With respect to that scheme, which is based on least mean squares localization and assumed the knowledge of the constant feature heights $^{C}p_2 =^{W} p_2 = h$, the solution adopted here introduces some improvements. We employ an EKF for estimation of feature coordinates and camera and environment parameters, maintaining only the assumption that the vehicle moves in a plane.

[??]

Knowing the current position of the feature $^{C}P_i$ in the camera frame (see equation (1)), related to $^{W}P_i$ in the fixed frame by a rigid-body motion (see fig. 2), and by some standard geometric calculations, a relationship among the feature coordinates in the fixed and moving frames with respect to the robot posture, is obtained as

$$^{C}\begin{bmatrix} p_1^i \\ p_3^i \end{bmatrix} = \begin{bmatrix} ^{W}p_3^i & ^{W}p_1^i & 1 & 0 \\ -^{W}p_1^i & ^{W}p_3^i & 0 & 1 \end{bmatrix} b, \qquad (2)$$

with

$$b = \begin{bmatrix} -\cos\xi_3 \\ \sin\xi_3 \\ \xi_2\cos\xi_3 - \xi_1\sin\xi_3 \\ -\xi_1\cos\xi_3 - \xi_2\sin\xi_3 \end{bmatrix} \qquad (3)$$

Equation (2) can be regarded as providing two nonlinear scalar equations in the 3 unknown robot state space variables (ξ_1, ξ_2, ξ_3), for each feature observed in the current and reference images. Therefore, localization problem is solvable for each position of the vehicle in $\mathbb{R}^2 \times S$ by solving – in a least-squares sense – for b, provided only that the $n \geq 4$ features do not belong to a single plane perpendicular to the plane of motion.

Strictly speaking, this approach works if camera measurements are noise-less, that is, in practical applications with low-cost visual apparatuses, unconsistent. Due to inverse perspective projection of 1 for the i–th feature

$$\begin{bmatrix} ^{C}p_1^i \\ ^{C}p_3^i \end{bmatrix} = \begin{bmatrix} \frac{\alpha_y}{\alpha_x} h_i \frac{x_i}{y_i} \\ \alpha_y \frac{h_i}{y_i} \end{bmatrix}, \qquad (4)$$

noisy image plane measurements are particularly dangerous as estimated features are close to the horizon, that is with y_i coordinate close to zero as shown in figure (5), corrupting the 3D feature estimation of $^{C}p_1^i$ and $^{C}p_3^i$ and the resulting localization. In [30] an heuristic algorithm named Feature Migration *has been proposed.* In this paper, an Extended Kalman filter has been adopted. Selecting a number of n feature for robot localization, estimated filter state is $S = [S_1, S_2, S_3, S_4, S_5, S_6, \ldots, S_{3n+1}, S_{3n+2}, S_{3n+3}]^T = [\xi_1, \xi_2, \xi_3, ^{W}p_1^1, ^{W}p_2^1, ^{W}p_3^1, \ldots, ^{W}p_1^n, ^{W}p_2^n, ^{W}p_3^n]^T$ where the fist three elements represent the vehicle localization. Unicycle kinematic model is assumed for state prediction, as

$$\begin{bmatrix} \hat{S}_1(k+1) \\ \hat{S}_2(k+1) \\ \hat{S}_3(k+1) \\ \hat{S}_4(k+1) \\ \hat{S}_5(k+1) \\ \hat{S}_6(k+1) \\ \vdots \\ \hat{S}_{3n+1}(k+1) \\ \hat{S}_{3n+2}(k+1) \\ \hat{S}_{3n+3}(k+1) \end{bmatrix} = \begin{bmatrix} S_1(k) + cos(S_3(k))u_1(k) + w_1(k) \\ S_2(k) + sin(S_3(k))u_1(k) + w_2(k) \\ S_3(k) + u_2(k) + w_3(k) \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{5}$$

where $U(k) = [u_1(k), u_2(k)]^T$ are the encoder measurements for forward and angular velocity, obtained from $u_1(k) = R\frac{\omega_r(k)+\omega_l(k)}{2}$ and $u_2(k) = R\frac{\omega_r(k)-\omega_l(k)}{L}$ respectively. ω_r and ω_l are the rotational encoder for the right and left wheel, R is the wheel radius and L is the length of the wheel axle. $w = (w_1, w_2, w_3)^T$ is the additive zero mean gaussian noise comprised of model dynamical errors and noisy odometry measurements. In a static structured environment, the feature are motionless in fixed frame $< W >$ and supposed to be known. The latter hypothesis may be relaxed including imperfect knowledge of feature position with respect fixed frame $< W >$. Knowing the localization estimation and the feature position in frame $< W >$, it is possible to obtain the feature coordinates with respect to mobile frame $< C >$ and, by inversion of equation 5, the estimated measurement to compare with CCD plane real coordinates. The measurement correction is, once again, added with zero mean gaussian noise. Covariance error matrices are weighted taking into account odometry lack of accuracy (typically due to wheels slipping and skidding) and [30] localization problems. Simulation results are reported in figure 6, with the least mean squares localization method and with the additional EKF localization compared for each vehicle coordinates.

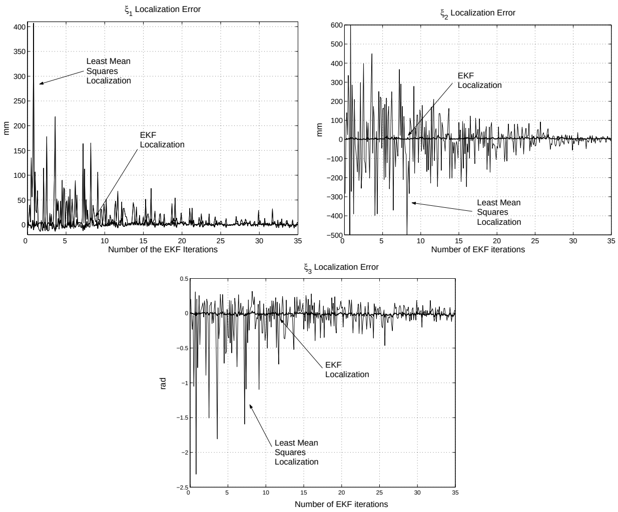

Fig. 6. Comparison between least mean squares and EKF localization methods for each robot coordinates $^W[\xi_1, \xi_2, \xi_3]^T$. As it was expected, the EKF acts as a noise low pass filter, improving localization noise rejection.

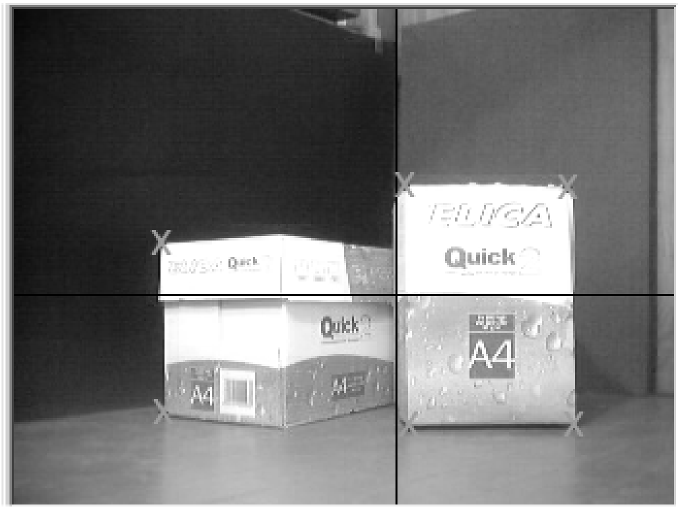

Fig. 7. Grabbed image. The image belongs to a certain topological map location, which corresponds to a 3D position $^W\bar{\xi}$.

IV. Virtual Visual Servoing

In a generic position the robot is localized using EKF described in section III, obtaining $^W\xi = {}^W[\xi_1, \xi_2, \xi_3]^T$ that corresponds to a specific image A (see fig. 3). Suppose that the robot has to reach a new position, say $^W\hat{\xi}$, expressed once again in the global metric reference map. If $^W\hat{\xi}$ corresponds to a topology mapped location, which has an associated image C, a graph visiting algorithm can be implemented to steer the vehicle so that $A \to C$, travelling through image B using the visual servoing controller. Note that a minimum travelling path algorithm could be implemented, constrained to topological map images, giving to each possible path between two images a proper weight. Closed loop connections between the path planners and the topological image map is ensured by the visual servoing controller.

If $^W\hat{\xi}$ has not yet been mapped in the topological map, the problem is now to reach an unknown, unmapped position. The first step is to retrieve information about the necessary visual servoing controller desired feature positions. Knowing mapped features position in the global metric map, it is now possible to reproject feature in the virtual mobile frame $< \hat{C} >$ positioned in the desired $^W\hat{\xi}$ by equation (2). Using equation (1), the feature in $< \hat{C} >$ are reprojected in the virtual image plane, using limited field-of-view camera constraint.

If a set α of at least four metrical mapped feature belongs to the virtual image plane, the topological image map furnish the image, say D, that has the same set α (as shown in figure 7, with the associated α set). At this point, the first problem is to steer the vehicle to image D, using, once again, the topological image map and a standard algorithm to visit the corresponding graph.

As the robot reaches the topological image D, that is the metric position $^W\bar{\xi}$, the transformation $^WF = [F_1, F_2, F_3]^T$

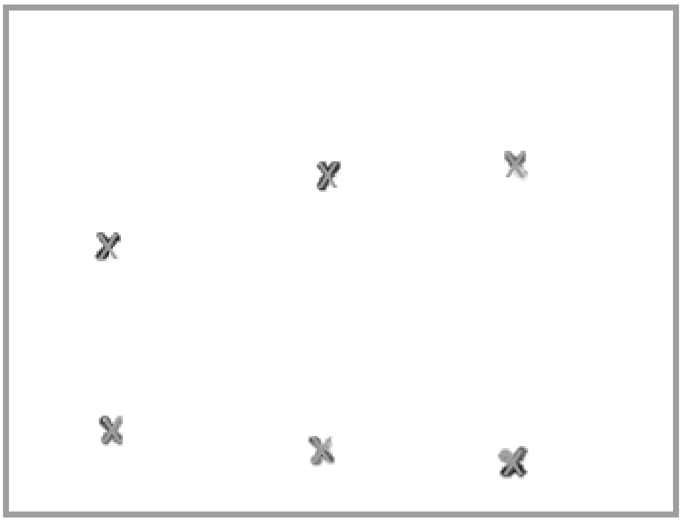

Fig. 8. Virtual image. The image contains reprojected image feature signed with crosses in fig. 7, as they appear in the unmapped desired position ${}^{W}\hat{\xi}$. The virtual image plane is obtained once the homogeneous transformation T_f has been computed.

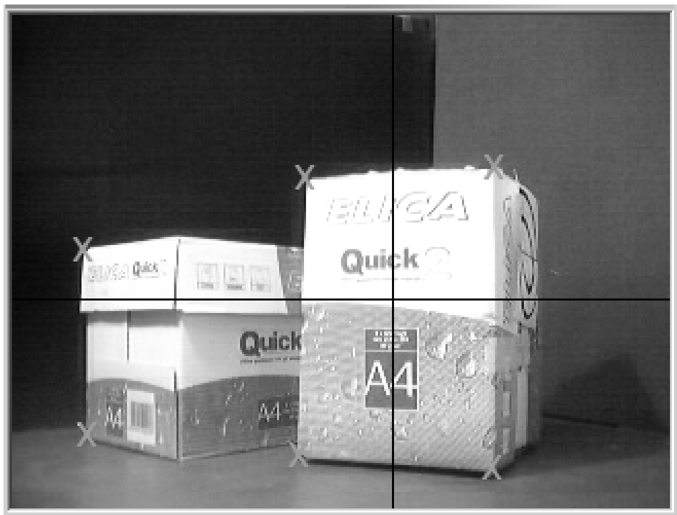

Fig. 9. Final servoed image. Note that image features correspond to the highlighted features in fig. 8, within the vision system noise.

between ${}^{W}\bar{\xi}$ and ${}^{W}\hat{\xi}$ is computed, simply noting that ${}^{W}F = {}^{W}\hat{\xi} - {}^{W}\bar{\xi}$. The homogeneous transformation $T_f = [R, t]$, where R is the rotation matrix of angle F_3 performed with respect to Y_C mobile frame axis and $t = [F_1, 0, F_2, 1]^T$ is the translation vector, is applied to the set α belonging to image D obtaining the virtual image (see fig. 8). The virtual image represents the desired, offset 3D feature position, expressed in the moving camera frame $< C >$. Virtual points in virtual image are used by the visual servoing controller as task targets.

It is worthwhile to note that outlier problem is intrinsically solved in the virtual image creation procedure. As aforementioned, the visual servoing controller works with the same paradigm and accuracy in real or virtual image, ensuring convergence in the desired, unmapped position. As the robot reaches the desired position and orientation ${}^{W}\hat{\xi}$, the real image is grabbed and added to topological map (see fig. 9), adjusting topological map arcs with corresponding feature information.

Generating virtual images, we have to deal with occlusion problems: given a desired location, the features of α are in the field of view of the virtual camera, but can be occluded by other objects during the servoed positioning. If the location of the projection belongs to a path between two mapped nodes of the topological map it is always possible to find reliable features without occlusions. It is worthwhile to note that servoing controller needs at least four features to complete the navigation, so occluded features can be discarded till the necessary minimum number is reached. In this work, the features in virtual image are simply chosen belonging to the closest path of the map to the virtual location, without going deeper in the occlusion problem. Furthermore, as the vehicle moves in the topological virtual images, it is able to identify and estimate new unknown features of the environment, growing the detail of the metric map and introducing new images in the topological map.

Existence of the set α of metrical mapped feature in the virtual image plane is a sufficient condition. If α does not exist, a heuristic position planner could be used, searching for a set of positions $\Gamma = \{\xi^j\}$, with $j = 1, \ldots, m$, where ξ^1 is a valid virtual position, i.e. where exist a set of valid mapped feature α^1, reachable from a topological image position. Both metric and topological maps, together with the Γ set, are built up as the robot navigates to the desired position, using desired global ${}^{W}\hat{\xi}$ position to drive the choices of subsequent ξ^j. If a solution exist, that is if the environment is sufficiently visual informative, last element of the Γ set, ξ^m and associated topological images, contains the necessary information to reach the desired position. Due to the "sufficiently visual informative" constraint, the described algorithm is heuristic and, at this time, there is not rigorous demonstration of its effectiveness.

It is worthwhile to note that within the proposed virtual images mechanism, the discrete topological image map could be treated as a continuous image stream of the environment, but storing only the finite number of necessary images.

V. Cooperation and Coordination

Cooperative localization and map building is another attractive research field, minimizing the time for building complex maps and incrementing the accuracy of the produced maps with augmented fault tolerance. Within the proposed solution, map building cooperation could be addressed once multiple agents start the exploration from known relative positions. Each vehicle built up its own maps with respect the fixed common frame. At a certain time, produced maps could be fused in a unique global map that could be shared among the autonomous vehicles.

As stated in section IV, using virtual images is possible to reach an unmapped position of the environment. This

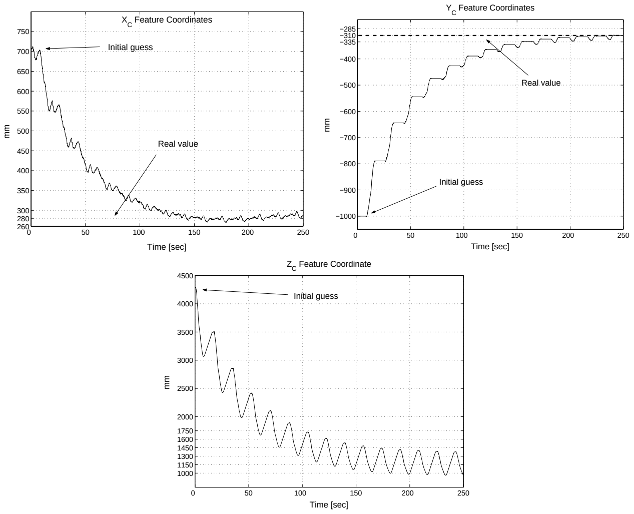

Fig. 10. Feature coordinates estimation with respect mobile camera frame $< C >$. Note that the Z_C coordinate varies depending on the particular controls applied to the robot.

potentiality is also used for robot cooperation, once common global maps are shared among the agents. Indeed, each robot is able to specify a global or relative target position for the rest of the robots, leaving to the latter the virtual images computation.

The particular implementation of the Visual SLAM for Servoing produce shared maps that are memory-safe and communication bandwidth-safe.

VI. Experimental results

A low-cost apparatus was employed, to highlight the applicability potential of the proposed technique. The experimental setup was comprised of a K-Team Koala vehicle [2], equipped with a cheap Kodak EZ200 web-cam [3] placed on the front part of the robot platform. The vehicle has two symmetric rows of three wheels on its sides, each actuated by a single low-resolution stepper-motor actuator: the construction implies that slipping and skidding of some of the wheels occurs whenever the vehicle moves along a curved trajectory. Such conditions make it hard to use odometry for localization and control, and strongly motivates the use of visual servoing. The controller is implemented under Windows XP on a 1130MHz Pentium III laptop mounted on-board. The Intel OpenCV [1] library wes used to compute optical flow and to track features. The hardware communication between the robot and the laptop is performed by a RS-232 serial cable.

Figure 10 shows experimental results for feature coordinates extimation. EKF techniques was employeed, starting with a strongly erroneous initial guess. The image based controller used in this first, preliminary experimental result was comprised of a simple back-and-forth controller, able to avoid feature occlusions and able to take into account limited field of view constraint.

Data retrieved from this experiment are relative to mobile camera frame $< C >$, so feature estimation for Z_C coordinate, that represents the distance of the feature from the robot, varies depending on image based controller that has been chosen.

VII. Conclusions

In this paper, we have proposed the Visual SLAM for Servoing, a preliminary solution to the well known SLAM problem based on a previously developed visual servoing control schema. The connection between control techniques (Action) and sensorial data interpretation (Perception) has been taken into account.

Economicity of the whole system has been addressed using off-the-shelf cameras fixed on a inexpensive consumer mobile platform.

The proposed solution is strictly related to the Extended Kalman Filter used to estimates feature positions.

Two dissimilar maps are naturally induced by the proposed architecture: a metric 3D maps, containing the estimated feature and robot posture as it is usually assumed in control literature, and a topological image map, composed of images of the mapped surrounding environment that an higher level of explored ambient perception.

Feature-based metric maps allows future improvements adopting more sophisticated visual feature extraction, like textures mapping or edge detection and tracking algorithms.

The proposed work implements a virtual visual servoing based on reprojected desired image feature, allowing autonomous exploration of unmapped environment regions.

The virtual visual servoing is also able to find a solution for coordination and cooperation among different autonomous agents. Produced maps are memory-safe and communication bandwidth-safe, facilitating the map sharing among mobile agents.

Simulations and preliminary experimental results on a laboratory vehicle are reported, showing the practicality of the proposed approach.

References

[1] *Intel Open Source Computer Vision Library*, http://www.intel.com/research/mrl/research/opencv/.

[2] *K-Team Koala Robot*, http://www.k-team.com/robots/koala/index.html.

[3] *Kodak EZ200 Web-Cam*, http://www.photo.net/ezshop/product?product_id=2371.

[4] P. Beeson, M. MacMahon, J. Modayil, J. Provost, F. Savelli, and B. Kuipers, *Exploiting local perceptual models for topological map-building*, in IJCAI-2003 Workshop on Reasoning with Uncertainty in Robotics (RUR-03), 2003.

[5] A. Bicchi, F. Lorussi, P. Murrieri, and V. G. Scordio, *MISTRAL - Methodologies and Integration of Subsystems and Technologies for Robotic Architectures and Locomotion*, STAR - Springer Tracts in Advanced Robotics, Springer-verlag, 2004, ch. On The Problem of Simultaneous Localization, Map Building, and Servoing of Autonomous Vehicles, pp. 223–239.

[6] J. Borenstein, H. Everett, and L. Feng, *Navigating mobile robots*, A.K. Peters, 1996.

[7] G. Campion, B. d'Andrèa Novel, and G. Bastin, *Controllability and state feedback stabilization of nonholonomic wheeled mechanical systems*, in Advanced Robotic Control, Springer-Verlag, 1991, pp. 106–124.

[8] G. Casalino, M. Aicardi, A. Bicchi, and A. Balestrino, *Closed loop steering and path following for unicycle–like vehicles: a simple lyapunov function based approach*, IEEE Robotics and Automation Magazine, 2 (1995), pp. 27–35.

[9] A. Comport, E. Marchand, and F. Chaumette, *Robust and real-time image-based tracking for markerless augmented reality*, Tech. Report 4847, INRIA, June 2003.

[10] F. Conticelli, D. Prattichizzo, F. Guidi, and A. Bicchi, *Vision-based dynamic estimation and set-point stabilization of nonholonomic vehicles*, in Proc. IEEE Int. Conf. on Robotics and Automation, San Francisco, CA, USA, 2000, pp. 2771–2776.

[11] B. d'Andréa Novel, G. Campion, and G. Bastin, *Control of nonholonomic wheeled mobile robots by state feedback linearization*, International Journal of Robotics Research, 14 (1995), pp. 543–559.

[12] A. Danesi, D. Fontanelli, P. Murrieri, V. G. Scordio, and A. Bicchi, *Cooperative simultaneous localization and map building for servoing*, in Proceedings of Multiagent Robotic Systems Trends and Industrial Application - Robocup 2003 Conference, Padova, July 2003.

[13] G. Dudek, M. Jenkin, E. Milios, and D. Wilkes, *Robotic exploration as graph construction*, in Proc. IEEE Int. Conf. on Robotics and Automation, vol. 7, December 1991, pp. 859–865.

[14] M. O. Franz, B. Schölkopf, P. Georg, H. A. Mallot, and H. H. Bülthoff, *Learning view graphs for robot navigation*, in Proc. of the First Int. Conf. on Autonomous Agents, W. L. Johnson and B. Hayes-Roth, eds., New York, 1997, ACM Press, pp. 138–147.

[15] J. Glasgow, N. H. Narayanan, and B. Chandrasekaran, *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, AAAI Press/The MIT Press, Cambridge, MA, 1995.

[16] K. Hashimoto and T. Noritsugu, *Visual servoing of nonholonomic cart*, in Proc. Int. Conf. on Robotics and Automation, April 1997, pp. 1719–1724.

[17] J. P. Hespanha, *Single camera visual servoing*, in Proc. IEEE Int. Conf. on Decision and Control, 2000, pp. 2533–2538.

[18] J. Kosecka, L. Zhou, P. Barber, and Z. Duric, *Qualitative image based localization in indoors environments*, in Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, Madison, Wisconsin, June 16–22 2003. To appear.

[19] B. Kuipers, *The 'map in the head' metaphor*, Journal of Environment and Behavior, 14 (1982), pp. 202–220.

[20] ———, *The spatial semantic hierarchy*, Journal of Artificial Intelligence, 119 (2000).

[21] B. Kuipers, D. Tecuci, and B. Stankiewicz, *The skeleton in the cognitive map: a computational and empirical exploration*, Journal of Environment and Behavior, 35 (2003), pp. 80–106.

[22] R. Kurazume and S. Nagata, *Cooperative positioning with multiple robots*, in Proc. IEEE Int. Conf. on Robotics and Automation, 1994, pp. 1250–1257.

[23] I. S. Kweon and T. Kanade, *High resolution terrain map from multiple sensor data*, IEEE Trans. on Pattern Recognition and Machine Intelligence, 14 (1992), pp. 278–292.

[24] J. J. Leonard, H. F. Durrant-White, and I. J. Cox, *Dynamic map building for an autonomous mobile robot*, International Journal of Robotics Research, 11 (1992), pp. 286–297.

[25] A. D. Luca, G. Oriolo, and M. Vendittelli, *Control of Wheeled Mobile Robots: An Experimental Overview*, vol. 270 of Lecture Notes in Control and Information Sciences, Springer Verlag, London, 2001, ch. RAMSETE. Articulated and Mobile Robots for SErvices and TEchnology, pp. 181–226.

[26] E. Marchand and F. Chaumette, *A new formulation for nonlinear camera calibration using virtual visual servoing*, Tech. Report 4096, INRIA, January 2001.

[27] ———, *Virtual visual servoing: a framework for real-time augmented reality*, in EUROGRAPHICS 2002, vol. 21, 2002.

[28] M. D. Marco, A. Garulli, S. Lacroix, and A. Vicino, *Set membership localization and mapping for autonomous navigation*, International Journal of Robust and NonLinear Control, 11 (2001), pp. 709–734.

[29] A. Matsikis, T. Zoumpoulidis, F. Broicher, and K. Kraiss, *Learning object-specific vision-based manipulation in virtual environments*, in Proceedings IEEE Int. Workshop on Robot and Human Interactive Communication ROMAN 2002, Berlin, September 2002, pp. 204–210.

[30] P. Murrieri, D. Fontanelli, and A. Bicchi, *A hybrid-control approach to the parking problem of a wheeled vehicle using limited view-angle visual feedback*, International Journal of Robotics Research, (2003). In press.

[31] D. Nieuwenhuisen and M. H. Overmars, *Motion planning for camera movements in virtual environments*, tech. report, Institute of Information and Computing Sciences,, Utrecht University, January 2003.

[32] J. Piaget and B. Inhelder, *The Child's Conception of Space*, Norton, New York, 1967. First published in French, 1948.

[33] I. Rekleitis, R. Sim, G. Dudek, and E. Milios, *Collaborative exploration for the construction of visual maps*, in Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, November 2001, pp. 1269–1274.

[34] S. I. Roumeliotis and G. Bekey, *Distributed multirobot localization*, IEEE Trans. on Robotics and Automation, 18 (2002), pp. 781–795.

[35] D. L. S. Se and J. Little, *Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks*, International Journal of Robotics Research, 21 (2002), pp. 735–758.

[36] S. Simhon and G. Dudek, *A global topological map formed by local metric maps*, in Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, October 1998, pp. 1708–1714.

[37] S. Thrun, *Robotic mapping: A survey*, in Exploring Artificial Intelligence in the New Millenium, G. Lakemeyer and B. Nebel, eds., Morgan Kaufmann, 2002.

[38] S. Thrun and A. Bücken, *Integrating grid-based and topological maps for mobile robot navigation*, in Proc. of the AAAI Thirteenth National Conf. on Artificial Intelligence, Portland, Oregon, 1996.

[39] S. Thrun, D. Fox, and W. Burgard, *A probabilistic approach to concurrent mapping and localization for mobile robots*, Machine Learning, 31 (1998), pp. 29–53. also appeared in Autonomous Robots 5, 253–271 (joint issue).

[40] S. Thrun, J.-S. Gutmann, D. Fox, W. Burgard, and B. Kuipers, *Integrating topological and metric maps for mobile robot navigation: A statistical approach*, in Proc. of the AAAI Fifteenth National Conf. on Artificial Intelligence, Madison, Wisconsin, 1998.

[41] D. Tsakiris, P. Rives, and C. Samson, *Applying visual servoing techniques to control nonholonomic mobile robots*, in International Conference on Intelligent Robots and Systems, September 1997. Workshop on "New Trends in Image-based Robot Servoing".

[42] S. Ullman, *High-level Vision: Object Recognition and Visual Cognition*, MIT Press, Cambridge MA, 1996.