A framework for visual landmark identification
based on projective and point-permutation invariant vectors

Christos I. Colios\textsuperscript{a}, Panos E. Trahanias\textsuperscript{*}

\textsuperscript{a} Institute of Computer Science, Foundation for Research and Technology—Hellas (FORTH),
PO Box 1385, Heraklion, 711 10 Crete, Greece
\textsuperscript{b} Department of Computer Science, University of Crete, PO Box 1470, Heraklion, 714 09 Crete, Greece

Received 23 December 1999
Communicated by F.C.A. Groen

Abstract

Autonomous robotic navigation based on machine vision, presents a challenging field in current robotics research. This task presupposes a representation of the environment and methods to process this representation and use it to derive the appropriate platform motions, in order to accomplish specific navigation goals. Following recent theories of active and purposive vision, we attempt to qualitatively describe the environment, avoiding thus a metric, 3D representation of it. A set of reference patterns, the so-called landmarks, together with their topological relations, is used to adequately describe the robot's workspace. Mathematical tools from projective geometry are employed for landmark identification, facilitating robust landmark recognition irrespective from the camera viewpoint. A complete framework is presented in this work for landmark extraction and recognition based on projective and point-permutation invariant vectors. Detailed experimentation has revealed accurate landmark recognition in indoor workspaces. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Landmark identification; Topological navigation; Invariant representations; Projective and permutation invariants

1. Introduction

Autonomous robot navigation using vision feedback presents a difficult research task due to the cognitive workspace representation it involves. Such a representation facilitates localization and the determination of the appropriate motion(s) to reach a navigation target. In many cases, cognitive representations of the environment manifest themselves as images of distinct objects or views of the workspace, the so-called landmarks, that can be used to unambiguously characterize the environment. The current paper addresses this issue employing results from projective geometry; it presents a framework for landmark learning during an initial, training phase, and landmark recognition during navigation.

\textsuperscript{*} Corresponding author. Tel.: +30-81-391-715; fax: +30-81-391-601.
E-mail addresses: colios@ics.forth.gr (C.I. Colios), trahania@ics.forth.gr (P.E. Trahanias).
The problem of autonomous navigation has been extensively studied in the past. In principle, two different approaches to this problem have emerged, metric and non-metric, based on the type of environment representation that is used. Metric approaches rely heavily on quantitative representations of the environment (environment maps). In contrast, non-metric approaches employ qualitative representations of the environment and hence result in qualitative navigation.

Qualitative environment representations involve, in most cases, a topological map that describes the workspace in terms of reference points and topological relations among these points. The set of reference points constitute the workspace landmarks that characterize segments of the environment as identifiable patterns. Visual landmarks are deemed to be salient features in the workspace, so as to be easily extracted and recognized from cluttered images taken in typical workspaces. Of major importance to such approaches is the landmark representation that is employed at the system level. Two broad classes of landmark representations are mainly used: iconic and non-iconic landmarks.

Perhaps one of the key approaches to non-iconic landmark representations is the fairly common use of straight lines [5]; see for a representative example the work by Kosaka et al. [13,14]. A variation of the above has been presented in [8] using panoramic images of the environment. Intensity values are averaged onto a one pixel wide circular band. Zero crossings of these 1D images typically correspond to wall corners, edges, or door frames, i.e. salient regions in images and therefore they can be selected as landmarks.

A simple way to define iconic landmarks is by employing entire images as such. Examples are the works presented in [6,21]. When the workspace is static and well defined in advance, either predesigned or selected landmarks can be employed, establishing in effect engineering approaches to landmark-based navigation [19,25]. It is evident that such approaches are inflexible and inherently confined to known and static environments.

In recent years many approaches to iconic landmark identification have focused on developing techniques for selecting salient areas in images, to delineate search regions for landmark patterns. This is usually done with the help of a saliency map [1,18,21,29], i.e. a data driven map of the image that identifies prominent landmark patterns. Such a bottom-up generation of attention cues can actually be combined with top-down, model driven search for known landmarks, as in the case of Trahanias et al. [28].

Landmark recognition during navigation is a difficult task due to pattern variations caused by perspective distortions. When an estimate of the robot position and orientation is available, the effect of perspective distortions can be removed facilitating thus landmark pattern comparisons [17,20]. In our previous work [28,29], we were able to perspective-correct the imaged patterns without the constraint of known robot position. This was achieved under the assumption of workspaces that form orthogonal parallelepipeds (e.g. hallways) and constraint egomotion of the observer. Other researchers have addressed this problem by imposing an appropriate platform motion [17,28].

In this paper, we propose a framework for landmark selection and recognition based on projective invariants. A 5D invariant vector is employed as the internal landmark representation. By exploiting the vector’s invariant nature we overcome the problem of landmark matching under different perspective projections, avoiding thus any specific assumptions about the workspace and/or the observer. The only assumption made regards the existence of planar surfaces, which is a prerequisite for the computation of the invariant vector. However, since we aim at indoor navigation, this assumption is valid in most such environments.

The proposed framework for landmark-based navigation consists of two distinct phases: the learning phase, being active during a training session and the recognition phase, being active during a navigation session. During the learning phase the robotic platform automatically identifies and stores sets of salient, coplanar quintuples as workspace landmarks, along with associated navigational information, building in effect a topological map of the environment. Landmarks are extracted and rediscovered at recognition time by means of projective and point-permutation invariant matching. The above operations are performed on the monocular, uncalibrated images of the viewed scene.

In what follows, Section 2 gives a brief introduction to results from projective geometry for the self-completeness of the paper. Section 3 presents the proposed method and in Section 4, we present experimental results. Finally, Section 5 concludes the paper with a brief discussion.
2. Projective and permutation invariant vector

Invariant descriptors seem to be a promising tool for representing workspace patterns since they enable recognition without the problem of pose estimation. The invariant vector we employ in this work is a combination of a projective and a point-permutation invariant; it has been studied in recent works by Lenz and Meer [16].

Let the 1D and 2D projective spaces \( \mathcal{P}^1 \) and \( \mathcal{P}^2 \), and let also the well-known projective invariant, the \textit{cross-ratio} \((\text{CR})\). In the case of \( \mathcal{P}^1 \), CR is calculated using four collinear points; let \( \text{CR}(x_0, x_1, x_2, x_3) = n \). There are \( 4! \) different point-permutations of \( x_i \)'s which could lead to the calculation of a cross-ratio, with only six of them giving rise to different values [24]. The calculation of the cross-ratio in \( \mathcal{P}^2 \) uses five coplanar points with no three of them being collinear; let it be \( \text{CR}(x_0, x_1, x_2, x_3, x_4) \). Similar to \( \mathcal{P}^1 \), the cross-ratio in \( \mathcal{P}^2 \) can be defined with other permutations of the five points. This means that there are \( 5! \) different permutations which have to be considered.

It follows from the above that direct application of the cross-ratio (either in \( \mathcal{P}^1 \) or in \( \mathcal{P}^2 \) for pattern representation and recognition purposes is not practical. In the case of \( \mathcal{P}^1 \) we need a vector of 24 components (\( 4! \)) to represent four points; the representation of five points in \( \mathcal{P}^2 \) becomes even more complicated since we should keep a vector with 120 (\( 5! \)) components. In either case the computational requirements would be considerably increased and, at the same time, the recognition accuracy would be reduced due to the limited resolution of the cross-ratio. A solution in both cases is to reduce the number of dimensions of these vectors, by using a point-permutation invariant.

Lenz and Meer [15] have derived permutation invariants for \( \mathcal{P}^1 \) and \( \mathcal{P}^2 \). In the current paper, we employ the permutation invariant \( J_{14} \) in \( \mathcal{P}^1 \) as a descriptor since (a) its value is bounded (between 2.0 and 2.8) and, therefore, does not suffer from instabilities due to singularities, and (b) as will be explained later, it can provide direct point to point correspondences across matched quintuples. It is defined as

\[
J_{14}(n) = \frac{2n^6 - 6n^5 + 9n^4 - 8n^3 + 9n^2 - 6n + 2}{n^6 - 3n^5 + 3n^4 - n^3 + 3n^2 - 3n + 1} .
\]

(1)

The invariant \( J_{14} \) is also suitable in \( \mathcal{P}^2 \). It can be shown [16] that a combination of the point-permutation in Eq. (1) with the cross-ratio yields to

\[
J_{14}(\text{CR}(x_0, x_1, x_2, x_3, x_4)) = J_{14}(\text{CR}(x_0, \pi(x_1, x_2, x_3, x_4))) ,
\]

(2)

where \( \pi(x_1, x_2, x_3, x_4) \) represents a permutation of the four points \( x_1, x_2, x_3, x_4 \). Notice that Eq. (2) is valid for all possible \( \pi \)'s and that any exchange of the first point with one of the last four points would violate (2). Therefore a projective and point-permutation invariant for a quintuple has to be a vector with five components.

In the case of \( \mathcal{P}^2 \) there exist only two independent cross-ratios [24]. Let them have the values \( m_0 \) and \( m_1 \). With these cross-ratio values and \( J_{14} \), a projective and point-permutation invariant vector (\( \mathcal{P} \mathcal{P} \mathcal{I} \mathcal{V} \)) can be derived as

\[
\mathcal{P} \mathcal{P} \mathcal{I} \mathcal{V} = \left( J_{14}(m_0), J_{14}(m_1), J_{14} \left( \frac{m_0}{m_1} \right), J_{14} \left( \frac{m_1 - 1}{m_0 - 1} \right), J_{14} \left( \frac{m_0(m_1 - 1)}{m_1(m_0 - 1)} \right) \right) .
\]

(3)

Each of the vector components in Eq. (3) is directly related to one of the five points: \( x_0, x_1, x_2, x_3, x_4 \). By ordering the components of \( \mathcal{P} \mathcal{P} \mathcal{I} \mathcal{V} \), e.g. in increasing order, direct point to point correspondences between two matched quintuples can be deduced, facilitating the application of additional constraints during quintuple matching.

---

\(^1\) An \( n \)-dimensional projective space \( \mathcal{P}^n \), can be thought of as arising from an \( (n+1) \)-dimensional vector space in which we define an equivalence relation between non-zero vectors; two such vectors \( x \) and \( y \) are equivalent if and only if \( y = \lambda x \). \text{CR} in \( \mathcal{P}^1 \) and \( \mathcal{P}^2 \) is an invariant quantity that intuitively expresses ratio of ratios of distances. For a more detailed treatment of projective geometry the reader is referred to [11].
3. Landmark Identification

As mentioned previously, this paper presents a framework for landmark-based navigation using PPIV of Eq. (3). A block diagram of this framework is shown in Fig. 1; as can be observed, the framework consists of two distinct phases:

- **Learning phase**: During an initial, training session, landmark patterns are detected and committed to memory along with associated navigational information, building in effect a topological map of the workspace.
- **Recognition phase**: During a navigation session, landmarks are recognized by successfully matching rediscovered patterns with stored ones. Recognition events facilitate platform localization and the determination of the next appropriate motion(s).

In the following sections, we present in detail these two phases. Since the issue of landmark representation is inherent in both phases, it is addressed first. Then we focus on landmark extraction and recognition, omitting details with respect to the topological map; this issue has been already studied by others [26].

3.1. Landmark representation

Visual landmarks constitute in the proposed framework non-iconic patterns that contain quintuples of points. Each quintuple is formed by coplanar points, with no three of them being collinear, and is regarded as a sub-landmark ($\mathcal{SL}$). A set (possibly with a single member) of sub-landmarks in an image frame forms a visual landmark ($\mathcal{VL}$). A formal definition of $\mathcal{VL}$ is as follows.

**Definition.** A visual landmark $\mathcal{VL}$ is a set of $k$ sub-landmarks $\mathcal{SL}_i$, $0 < i \leq k - 1$, which are sets of five coplanar points, with no three of them being collinear.
3.2. Learning phase

During training, potential landmarks are extracted, examined and, if appropriate, committed to memory for future reference. As a first step, \( SL \) candidates are detected using a set of salient points on the image. Such points are extracted as prominent corners in the image that reside in salient areas, and their detection is performed purely bottom-up (data driven) in order to avoid any specific assumptions about the environment.

3.2.1. Extraction of salient points

In our case the features of interest are image points and their robust extraction is the first step towards effective landmark identification. There are various ways that image points can be defined, such as curvature maxima, intensity extrema, corners, saddle points, etc. In this work we deal mostly with indoor workspaces, where geometric structures are usually observed in the background of large flat areas. Corners tend to be salient points in such scenes and, moreover, their robust and consistent detection is a tractable task. Therefore, they have been employed as the points of interest in the proposed framework.

In order to extract landmark patterns that would facilitate their own robust recognition, we employ a focus of attention mechanism on salient image areas. Such techniques have also been used by others for landmark selection [7]. The attention mechanism we employ is the saliency map. This represents a map of the image where distinctive areas are given larger values and areas that correspond to smooth regions are given smaller values [7, 18]. In the current implementation we have employed the saliency map proposed by Trahanias et al. [28] due to its suitability to indoor workspaces.

Having detected a salient area, the points of interest (corners) within it should next be extracted. Since we are interested in robust and consistent detection of a fairly small number of corners in a salient area, a criterion of goodness is particularly important for each extracted corner. The KLT-corner detector (KLT-CD) [22, 27] operates based on such a criterion; consequently, it has been employed as the point extractor in our approach. KLT-CD scans the salient area in the incoming image and, robustly, selects the \( N \) best corners according to the goodness criterion [4]. The latter is determined by examining the smallest eigenvalue of a \( 2 \times 2 \) gradient matrix centered at the image points in the area under consideration. Its robustness stems from the fact that it tracks the detected features in the next frame before reporting their detection. Moreover, each corner is explicitly tracked in a number of frames (four in our implementation) in order to detect and reject corners that do not appear consistently over successive image frames.

3.2.2. Selection of sub-landmarks

The next step in the learning phase is the formulation of legal quintuples, i.e. quintuples of coplanar points, with no three of them collinear, since (Section 2) the cross-ratio is preserved under projective transformations only if the five points used for its calculation form a legal quintuple. These legal quintuples are then considered as sub-landmark (\( SL \)) candidates; identified \( SLs \) are used in the sequel for building the visual landmarks (\( VLs \)). Theoretically, we could form quintuples from corners that are located in different salient areas. These would yield \( SLs \) with very wide topological dispersion causing, in effect, additional obstacles to landmark recognition. Thus, we confine the search for legal quintuples to corners from the same salient area, as identified by the saliency map.

Each quintuple located in a salient area under consideration is tested to verify that: (a) it does not contain triplets of collinear points, and (b) it contains coplanar points. The restriction of non-collinearity is directly related to the calculation of the cross-ratio, since three collinear points yield a determinant equal to zero and in effect a value of the cross-ratio equal to zero or \( \infty \). The restriction of coplanarity stems from the requirement of invariance of the cross-ratio and, consequently, of \( \mathcal{P} \mathcal{P} \mathcal{I} \mathcal{V} \).

\(^2\) The term consistent denotes uniform detection of the same world points as corners in different frames of the same scene.
3.2.2.1. Point collinearity. Although collinearity is preserved under perspective transformations, quasi-collinearity is not. This means that three quasi-collinear points in one frame could be collinear in another one and, therefore, quasi-collinearity has to be taken into account. A robust criterion for testing against collinearity or quasi-collinearity can be formulated by employing the **Grim matrix**, $M_{012} = \sum_{i=0}^{2} x_i x_i^T$ [12]. The smallest eigenvalue of $M_{012}$ is also the smallest eigenvalue of the matrix $\lambda_{012} = (x_0 x_1 x_2)$, i.e. the matrix having as columns the vectors of the three points [3]. The closeness to rank deficiency of $\lambda_{012}$ (the closeness of its smallest eigenvalue to zero) is a robust measure of the collinearity of the three points. A threshold $t_{cl}$ has been experimentally obtained and used in our implementation to test the closeness to rank deficiency of $\lambda_{012}$; if the smallest eigenvalue of $\lambda_{012}$ is less than $t_{cl}$ the points are considered as (quasi)collinear and the corresponding quintuple is excluded from the list of legal quintuples.

3.2.2.2. Point coplanarity. Quintuples passing the collinearity test are then examined to verify the coplanarity of their constituent points. This is achieved through the invariance of $\mathcal{PPIV}$ in corresponding quintuples in different image frames. Unfortunately, the invariance criterion is only a strong indication and not necessarily a proof of coplanarity. Moreover, $\mathcal{PPIV}$ suffers from numerical instabilities and cannot be used alone as a robust criterion for coplanarity. In the proposed framework it is augmented with two additional criteria: the **convex hull test** and the **projectivity test**.

**$\mathcal{PPIV}$ invariance.** A legal quintuple observed in two different frames results in the same value of $\mathcal{PPIV}$ (Eq. (3)). This can be used as a test for coplanarity, by examining the condition

$$\mathcal{PPIV} - \mathcal{PPIV}' = 0,$$

(4)

where $\mathcal{PPIV}$ and $\mathcal{PPIV}'$ are the calculated invariant vectors in the two image frames. However, the projective invariant $J_{14}$ (Eq. (1)) is bounded between 2.0 and 2.8. Numerical instabilities due to noise or small distortions in corner detections, combined with the above-mentioned range, pose serious obstacles in the strict application of Eq. (4), and the notion of constancy is modified to

$$|\mathcal{PPIV} - \mathcal{PPIV}'| \leq t_{\mathcal{PPIV}},$$

(5)

where $t_{\mathcal{PPIV}}$ is a predefined threshold value. An important issue in the coplanarity test of a quintuple observed in frame $F_k$, is the establishment of a search area in frame $F_{k+j}$, where potential matching quintuples are located. Let $A$ be the area defined by the smallest circumscribing rectangle of the quintuple in $F_k$. In $F_{k+j}$, we focus on $A$ expanded appropriately to cater for the motion of the robot, which is assumed smooth. In our implementation we choose the two frames adjacent in time; in this way we apply a small expansion to $A$, confining thus the search for matching quintuples in a limited area.

**Convex hull test.** The convex hull of a point set is preserved under projective transformations [9,10]. This, combined with the fact that a pair of matched quintuples in two different frames puts also their constituent five points in correspondence, provides the ability to use the **convex hull invariance** as an additional matching criterion. This invariance is expressed with the following conditions: (a) preservation of the number of points on the convex hull, (b) corresponding points lie either on or inside the convex hull, and (c) for points lying on the convex hull, neighboring relations are preserved. The last condition is useful if at least four points lie on the convex hull and, therefore, we accept only quintuples with four or five points on the convex hull.

Let a quintuple in frame $F$ and $n$ quintuples in frame $F'$, represented with the sets of points $x_i$ and $x'_j$, respectively, with $0 \leq i \leq 4$ and $0 \leq j \leq n - 1$. Let also the convex hull of the quintuple in $F$ be $\mathcal{CH}$ and the convex hulls of the quintuples in $F'$ be $\mathcal{CH}'_j$. We assume that the calculated vectors $\mathcal{PPIV}$ and $\mathcal{PPIV}'_j$ satisfy Eq. (5) $\forall j \in [0, n]$. In other words there are more than one matching candidates in $F'$ for the quintuple in $F$.

The fact that the five components of $\mathcal{PPIV}$ suffer from numerical instabilities may introduce mismatches in providing point to point correspondences. Therefore, we employ the convex hull test as a criterion to prune false matching candidates, and also to detect and correct erroneous point correspondences. That is, when two false point
to point correspondences are detected between the quintuple in $\mathcal{F}$ and a matching candidate with index $j$ in $\mathcal{F}'$ then we assume that these mismatches could be possibly due to numerical instabilities in the calculation of $\mathcal{PPIV}$ and $\mathcal{PPIV}'$. For example $\mathcal{PPIV} \text{ and } \mathcal{PPIV}'_j$ could yield the following point correspondences:

$$\{x_0, x'_3\}, \{x_1, x'_0\}, \{x_2, x'_4\}, \{x_3, x'_2\}, \{x_4, x'_1\}. \quad (6)$$

The convex hull test may detect that two point pairs, e.g. the first and the last one, are false. Then quintuple $j$ will not be excluded from the list of possible matching candidates; on the contrary it will be temporarily assumed that these point pairs have been mismatched due to inaccuracies in the calculation of the vectors $\mathcal{PPIV}$ and $\mathcal{PPIV}'_j$ and they will be mutually exchanged. Thus the corrected point pairs would be

$$\{x_0, x'_{1j}\}, \{x_1, x'_{0j}\}, \{x_2, x'_{4j}\}, \{x_3, x'_{2j}\}, \{x_4, x'_{3j}\}. \quad (7)$$

If more than two false point pairs are detected, then the quintuple in $\mathcal{F}'$ is rejected since there is no additional information about the point pairs. In conclusion, the application of the convex hull test has a twofold result: (a) false matching quintuples in $\mathcal{F}'$ are detected and rejected, and (b) erroneous point correspondences across quintuples in $\mathcal{F}$ and $\mathcal{F}'$ are temporarily inverted (corrected), and the corresponding quintuples in $\mathcal{F}'$ are retained. The final selection of the corresponding quintuple in $\mathcal{F}'$ is obtained by applying the projectivity test to all retained quintuples.

Projectivity test. Let the set $x'_0, 0 \leq i \leq 4, 0 \leq j \leq n - 1$, of retained quintuples in $\mathcal{F}'$, that are matching candidates of the quintuple $x_i$ in $\mathcal{F}$. The projectivity test is applied in this step in order to detect the quintuple, if one exists, in $\mathcal{F}'$ for which the best point to point correspondences between $\mathcal{F}$ and $\mathcal{F}'$ exist. The test is based on the fact that the transformation that maps the quintuple in $\mathcal{F}$ with the actual corresponding quintuple in $\mathcal{F}'$ should be a projective transformation ($3 \times 3$ non-singular matrix that is defined up to a scale factor). Four known point to point correspondences can be used to estimate this transformation (matrix); the actual estimation is derived from the solution of a linear system of eight equations with the same number of unknowns [11, p. 483]. The four points used constitute a projective basis; there are five different projective bases since we have five known point to point correspondences. A projective basis is used to calculate the projective transformation, which is then employed to estimate the position of the fifth point. The Euclidean distance of the estimated point position from its actual position is used to quantify the accuracy of the point correspondences.\textsuperscript{3} These steps are repeated for all five projective bases. The average of the five calculated distances is taken as the result of the projectivity test, referring to the quintuple in $\mathcal{F}$ and to one of the $n$ possible matching candidates in $\mathcal{F}'$.

The result of the projectivity test is subsequently used in order to identify one quintuple in $\mathcal{F}'$ as corresponding to the one in $\mathcal{F}$, or to declare the non-existence of a corresponding quintuple in $\mathcal{F}'$. The quintuple in $\mathcal{F}'$ with the lowest projectivity test value is identified as the correct matching quintuple, provided that this value is below a predefined threshold value $t_p$. In the opposite case, we conclude that no quintuple in $\mathcal{F}'$ corresponds to the quintuple in $\mathcal{F}$.

3.2.3. Visual landmark construction

In the proposed framework, a $VL$ consists of a set of $SL$s. In order to minimize the effect of numerical instabilities at recognition time, we select, from all detected, legal $SL$s, as most promising candidates the outliers. Vector ordering techniques [2], based on the Euclidean distance from the mean invariant vector, are used to identify such outliers.

Let $q$ quintuples identified as $SL$ candidates. In order to select the outliers we proceed as follows. At first the mean invariant vector $\mathcal{PPIV}_{\text{mean}}$ is calculated, $\mathcal{PPIV}_{\text{mean}} = (1/q) \sum_{i=1}^{q} \mathcal{PPIV}_i$, where $\mathcal{PPIV}_i$ is the vector of the $i$th candidate $SL$. Next, the outliers are selected based on the distances of candidate $SL$s from $\mathcal{PPIV}_{\text{mean}}$.

\textsuperscript{3} If a projective transformation between two views is calculated from the positions of only four points then it is vulnerable to noise in the data [23]. In our case, however, the estimated transformation is verified through its application to the fifth point. Therefore, transformations that are estimated from noisy data are most likely excluded from further consideration.
given as

\[ d_i = \sqrt{\sum_{j=0}^{4} (PPIV_i^j - PPIV_{mean})^2}, \]

where the superscript \( j \) denotes the \( j \)th vector component. The most outlier \( SL_i \)'s are detected as the ones with \( d_i \) greater than twice the average of all \( d_i \)'s. If no such \( SL_i \)'s exist, then no \( VL \) is reported in this particular case. Otherwise, the set of selected \( SL \)'s constitutes the \( VL \). This is stored along with attributes (\( PPIV \) vectors and convex hulls) of its constituent \( SL \)'s, that need to be retained for future reference.

3.3. Recognition phase

During a navigation session landmark recognition facilitates self-localization and the determination of the next platform motions. In this phase, a process, similar to that used in the learning phase, is employed to trace legal quintuples. Next we attempt to match each detected coplanar quintuple to one sub-landmark of a stored visual landmark, i.e. we try to build a visual landmark hypothesis.

In order to find the most promising stored candidate for each quintuple we calculate the Euclidean distance of the latter’s invariant vector from the vectors of every stored sub-landmark. Starting from the most promising stored sub-landmarks, we search for matching candidates that stand the convex hull and the projectivity constraints. The matched sub-landmark is used thereon as a guide for the detection of other sub-landmarks that will strengthen our landmark hypothesis. Once the hypothesis is verified to a certain extend, the landmark is considered recognized. In case of unsuccessful matching, the extracted coplanar quintuples can be used for building new landmarks, i.e. for a constrained update of the topological map.

4. Experimental results

The proposed framework has been implemented and experimentally verified in an indoor environment, namely the corridors outside our laboratory. TALOS, a mobile robotic platform available in the laboratory has been used as a testbed in our experiments. In this section, we present sample results from these experiments; a detailed illustrative result is first given to demonstrate the learning and recognition phases. Additional sample results and quantitative results are then presented for evaluation purposes.

In these experiments the threshold values have been set experimentally using image sequences that have been acquired by TALOS in the mentioned workspace. The criterion used to set the threshold values was conformance of the obtained landmark identification results, with manually obtained results on a number of image frames. The operational values used in our implementation are: \( t_{cl} = 0.001, t_{PPIV} = 0.08 \). Threshold \( t_p \) is given different values, according to the task being active; during the extraction of legal quintuples it is assigned a value of \( 1.4 \), whereas in quintuple matching (recognition phase) its value is equal to 12. The discrepancy in the values of \( t_p \) is justified by the fact that extraction of legal quintuples is performed by considering neighboring frames, resulting in small values in the result of the projectivity test; quintuple matching (recognition) usually refers to distant frames, with obvious effect in the result of the projectivity test value. More importantly, this experimentation has revealed that the threshold values are not critical for the overall system performance; moderate variations in their values had practically no effect in the observed performance of our implementation.

4.1. Illustrative results

Several experiments have been conducted to test the proposed framework. In this section, we present sample results for illustrative purposes. The first result is presented in detail and refers to the workspace scene that is
observed when the mobile platform reaches the end of a corridor, which forms a “T” junction with another corridor (Fig. 2).

4.1.1. Learning phase

A sequence of five consecutive frames encountered at learning phase is shown in Fig. 2, where the corners extracted by KLT-CD on each frame independently are superimposed on the corresponding frame. Detection of the consistent corners, i.e. corners extracted in the first frame with subsequent tracking of their persistence in the next frames, has resulted in 27 corners which are depicted in Fig. 3(a). By consulting the saliency map, an area of interest (the poster area on the right wall) has been identified and the corner points contained in it have been isolated for further investigation; 17 corners were identified, which are shown in Fig. 3(b).

![Fig. 2. Sequence of five frames encountered at learning phase.](image)

![Fig. 3. (a) Corners extracted at learning phase; (b) subset of corners in (a) residing in the salient area.](image)
From the 6188 possible quintuples that can be formed using the 17 corners, the application of the collinearity test (quintuples with no three points being collinear) has resulted in 478 quintuples. This set is further pruned by accepting only quintuples with 4 or 5 points on their convex hull resulting in 100 quintuples. These are tested for coplanarity by verifying constancy of the invariant vectors. To further strengthen our hypothesis we applied the convex hull and the projectivity test. A set of 17 quintuples were identified as coplanar and also satisfied the above-mentioned
criteria. Thus, these quintuples represent the strongest features in the salient area under consideration. From the set of 17 quintuples, the three most outliers according to Eq. (8) were finally selected as sub-landmarks, referred here as $S \mathcal{L}_0$, $S \mathcal{L}_1$ and $S \mathcal{L}_2$, respectively. Their graphical representation, along with the corresponding values of the invariant vectors $P P I V_i$, $0 \leq i \leq 2$, is given in Fig. 4. These three sub-landmarks constitute a visual landmark, which is inserted, along with navigational information, in the topological map for future reference.

4.1.2. Recognition phase

During navigation the same scene is viewed from a different vantage-point (Fig. 5). Under a similar procedure (Section 3.3) coplanar quintuples are extracted and successful matches with stored patterns are examined. From the detected coplanar quintuples we were able to successfully match three of them, namely $S \mathcal{L}_0$, $S \mathcal{L}_1$ and $S \mathcal{L}_2$, with the three stored sub-landmark patterns $S \mathcal{L}_0$, $S \mathcal{L}_1$ and $S \mathcal{L}_2$, respectively. These quintuples constitute the recognized sub-landmarks, and are shown in Fig. 6.

An interesting observation regards the recognition of sub-landmark $S \mathcal{L}_0$. Due to numerical instabilities in the calculation of $P P I V_0$, two erroneous point correspondences have resulted. However, the application of the convex-hull test has detected and appropriately corrected this mismatch. The mentioned points are the ones depicted with the horizontal arrows in Fig. 6 ($S \mathcal{L}_0$). In the other two cases the invariant vector provided the correct point to point correspondences. The projectivity test resulted in the values 3.65, 10.24 and 1.73 for the three sub-landmarks $S \mathcal{L}_0$, $S \mathcal{L}_1$ and $S \mathcal{L}_2$, respectively.

To further demonstrate the performance of the proposed framework, Fig. 7(a)–(c) shows three additional $S \mathcal{L}$ recognition results obtained. In each of these results, the left image shows the $S \mathcal{L}$ identified at learning phase, whereas the right image presents the $S \mathcal{L}$ recognition result during navigation.

4.2. Quantitative results

In order to quantitatively evaluate the proposed framework, we have set up an experiment in the workspace of the first illustrative result. The evaluation scenario we considered consisted in 50 navigation trials; each time sub-landmark recognition events have been reported. The results obtained from this experiment are summarized in Table 1. Each row in the table indicates, for each sub-landmark, the number and percentage of: (a) correct recognitions (CR), (b) mis-recognitions (MR), i.e. when the sub-landmark $S \mathcal{L}_i$ was reported as $S \mathcal{L}_j$, (c) false positives (FP), i.e. when a sub-landmark was erroneously reported, and (d) false negatives (FN), i.e. when a sub-landmark was missed.

As can be observed from Table 1, the proposed framework has resulted in a 93.3% correct recognition rate. By investigating the cases where errors have occurred, we concluded that these were mainly due to: (a) inaccuracies in the detection of the position of the corners, (b) extreme placement of the camera towards the wall containing the
Fig. 6. Recognized sub-landmarks.

The sub-landmarks, and (c) small variations of the salient areas as detected by the saliency map. The above manifest themselves as either errors in the calculation of various parameters involved, distortions in the landmark patterns, and false positive/negative alarms with respect to the presence of a (sub-)landmark. The obtained recognition accuracy of 93.3% is a promising result that may be further improved by considering previous (sub-)landmark recognition events during the current recognition task.
Fig. 7. Sample SL identification results.
Table 1
Landmark recognition results

<table>
<thead>
<tr>
<th>SL</th>
<th>CR</th>
<th>MR</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL₀</td>
<td>47</td>
<td>94</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>41</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>SL₁</td>
<td>44</td>
<td>88</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>23</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>SL₂</td>
<td>49</td>
<td>98</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>36</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>46.7</td>
<td>93.3</td>
<td>0.7</td>
<td>1.4</td>
</tr>
</tbody>
</table>

5. Discussion

In this paper, a framework has been presented for automated landmark identification, to facilitate autonomous robotic navigation. Projective and point-permutation invariant vectors have been employed to characterize landmark patterns. The use of these vectors relies on the existence of planar features in the workspace, which is valid in most indoor workspaces. The invariant vectors establish direct point to point correspondences, which enables the use of the convex hull and projectivity constraints. These constraints minimize the risk of mismatches across quintuples in different image frames. Moreover, the final selection of sub-landmarks as outlier patterns facilitates robust landmark recognition during navigation. Experimentation with this approach has revealed its effectiveness for a realistic indoor environment.

Future work is planned in the efficient implementation of the proposed framework as well as its enhancement in order to consult the topological map and previous (sub-)landmark recognitions to confine pattern matching against only a few candidate quintuples. Another direction of study concerns the use of geometric invariants combined with other invariants, such as photo-invariants, for pattern recognition. Additionally, uncertainty issues due to total or partial occlusions of stored landmarks at recognition time need further investigation.

References


Christos I. Colios has graduated in 1998 from the University of Hannover, Hannover, Germany, where he studied electrical engineering. He pursued his final year thesis in the subject of autonomous robotic navigation using visual workspace landmarks in the University of Crete, Heraklion, Greece. After finishing his studies he assumed a position as Research Assistant in the Computer Vision and Robotics Laboratory, Institute of Computer Science, Foundation for Research and Technology — Hellas (FORTH), Heraklion, Greece. His research interests lie in the areas of digital signal and image processing, robotic vision and computer aided scene analysis.

Panos E. Trahanias is an Associate Professor with the Department of Computer Science, University of Crete, Greece, and the Institute of Computer Science, Foundation for Research and Technology — Hellas (FORTH), Greece. He received his Ph.D. in Computer Science from the National Technical University of Athens, Greece, in 1988. He has been a Research Associate at the Institute of Informatics and Telecommunications, National Center for Scientific Research “Demokritos”, Athens, Greece. From 1991 to 1993, he was with the Department of Electrical and Computer Engineering, University of Toronto, Canada, as a Research Associate. He has participated in many RTD programs in image analysis at the University of Toronto and has been a consultant to SPAR Aerospace Ltd., Toronto. Since 1993, he is with the University of Crete and FORTH; currently, he is the supervisor of the Computer Vision and Robotics Laboratory at FORTH, where he is engaged in research and RTD programs in vision-based robot navigation. He has published over 60 papers in technical journals and conference proceedings and has contributed in two books.