Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 870492, 15 pages
doi:10.1155/2008/870492

Research Article

Global Interior Robot Localisation by a Colour Content Image Retrieval System

A. Chaari,[1,2] S. Lelandais,[1] C. Montagne,[1] and M. Ben Ahmed[2]

[1] *IBISC Laboratory, CNRS FRE 2873, University of Evry 40, Rue du Pelvoux, 91020 Evry Cedex, France*
[2] *RIADI Laboratory, National School of Computer Science, University of Manouba, 2010 La Manouba, Tunisia*

Correspondence should be addressed to A. Chaari, anis.chaari@ibisc.fr

Received 2 October 2006; Revised 10 April 2007; Accepted 3 August 2007

Recommended by Jose C. M. Bermudez

We propose a new global localisation approach to determine a coarse position of a mobile robot in structured indoor space using colour-based image retrieval techniques. We use an original method of colour quantisation based on the baker's transformation to extract a two-dimensional colour pallet combining as well space and vicinity-related information as colourimetric aspect of the original image. We conceive several retrieving approaches bringing to a specific similarity measure D integrating the space organisation of colours in the pallet. The baker's transformation provides a quantisation of the image into a space where colours that are nearby in the original space are also nearby in the output space, thereby providing dimensionality reduction and invariance to minor changes in the image. Whereas the distance D provides for partial invariance to translation, sight point small changes, and scale factor. In addition to this study, we developed a hierarchical search module based on the logic classification of images following rooms. This hierarchical module reduces the searching indoor space and ensures an improvement of our system performances. Results are then compared with those brought by colour histograms provided with several similarity measures. In this paper, we focus on colour-based features to describe indoor images. A finalised system must obviously integrate other type of signature like shape and texture.

1. INTRODUCTION

The autonomous robot navigation in a structured interior or unstructured external environment requires the integration of much functionality, which goes from the navigation control to the mission supervision, while passing by the perceived environment modeling and the planning of trajectories and strategies of motion [1]. Among these various functionalities, the robot localisation, that is, the capacity to estimate constantly its position is very significant. Indeed, the knowledge of the robot position is essential to the correction of trajectory and the execution of planned tasks.

Sensors constitute the fundamental elements of a localisation system. According to the type of localisation needed, we can use either proprioceptive sensors or exteroceptive sensors. Proprioceptive sensors measure displacements of the robot between two moments. The integration of their measures allows estimating the current position of the robot compared to its starting one. On the other hand, the exteroceptive sensors measure the absolute position of the robot by observing benchmarks whose positions are known in an environment frame-attached reference.

The localisation problem is fundamental in mobile robotics and always pokes a crescent number of contributions. DeSouza and Kak propose in [2] an outline of the various approaches, as well in interior structured as in external unstructured environments. These techniques can be gathered in two principal categories: relative localisation methods and absolute localisation methods:

(i) relative or incremental localisation where the robot position is computed by incrementing its preceding position and the measured variation with proprioceptive sensors (the two principal methods of relative localisation are odometry and the inertial localisation, these techniques use unstructured data and produce an accumulating error to estimate the robot position);

(ii) absolute localisation requires the knowledge of the environment to determine exactly the robot position or

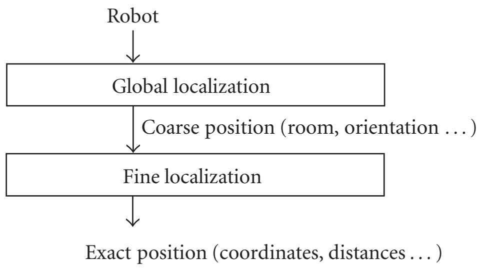


Figure 1: Proposed global localisation task which aims to give a coarse position of the robot. These global localisation's outputs could be used to keep only a part of the indoor space as inputs to a fine and exact localisation system for navigation purpose.

to periodically readjust incremental estimate (navigation) produced with relative localisation techniques. Exteroceptive sensors are used and various techniques can be distinguished to compute the robot position. The most known approaches are the magnetic compasses localisation, the active reference marks localisation, the passive reference marks localisation, and the model-based localisation techniques [3].

We propose in this paper a new approach for the robot localisation problem which consists in using an image database model and consequently content-based image retrieval techniques to provide a qualitative and a coarse estimate of the robot position. The central idea is to provide to the system a set of images and features potentially visible and detectable by computer vision techniques. The system's aim, thus, consists in searching attributes and features to identify the closest images from this set which indicate a coarse position and orientation of the robot. We introduce thus the term of global localisation which aims to indicate a coarse position of the robot like its room or orientation and which is different from fine or absolute localisation problem. This global localisation generally intervenes before the fine localisation process which aims to compute accurately the robot position (cf. Figure 1). We intend by fine localisation any localisation system developed for a purpose of robot navigation and which gives an exact position of the robot. The next section gives an overview of this fine localisation systems which could be as well map-based systems as maples systems.

In this work, we developed a global localisation robotic solution for disabled people within a private indoor environment. This global localisation could simplify the fine localisation by searching the robot position in a simple part of the space instead of the entire environment. Moreover, this global localisation is necessary after a long displacement of the robot to know its position whether it is lost and when the problem of fine localisation is difficult to solve.

We work through the ARPH project (Robotics Assistance to Handicapped People) [4] defined with the French Association against Myopathies (AFM). The aim of the project is to embark an arm manipulator (see Figure 2) on an autonomous mobile basis. By using the arm, a handicapped person is able to carry out various tasks of the current life. The various control modes include or not the handicapped person. Thus, the base must be able to be completely autonomous. To ensure this capacity, various sensors equip the base: proprioceptive odometric sensors for the relative localisation, ultrasonic sensors for obstacles detection and a colour camera as exteroceptive sensors.

Figure 2: Prototype of the handicapped person assistance's robot.

For the global localisation, we use the colour camera fixed in the base and we propose a content-based image retrieval method. The principle is to build an image database of the indoor space in which the robot evolves/moves. To find itself, the robot takes an image of its environment called request image. Then the system seeks the closest image from the database to the request image from which it deduces the room and the orientation of the robot.

Unlike most retrieval systems, request images taken by the robot's camera differ from images stored in the database. Although, the image database describes the totality of the indoor environment, the random navigation of the robot (according to the implicit need of the handicapped person) always gives different request images from those of the database. It is a question of extracting from the database, the closest image compared to the request image. This image will be used to determine the room where the robot is and its orientation in this room: two essential information needed for the global localisation of the robot in an indoor environment. In order to achieve this goal, colour information is needed. Unfortunately, illumination is not controlled and is not known to have invariant template against its changes. In addition, many small objects are removable and make partial occlusion of other objects. Thus it is necessary to rather seek features which tolerate these changes, from which one can find the image in question, than nonstable and complete features, which proves too restrictive. What is required is the compactness of features with the rapidity of computation since the image database is not very bulky.

The remainder of this paper is organised as follows. In the next section, we present related works on indoor robot localisation and content based image retrieval systems. Data we used is presented in Section 3. In Section 4, we develop the colour histograms techniques for image retrieval systems. The components and details of our retrieval system are described in Sections 5 and 6, respectively. We present and discuss our results in Sections 7 and 8 and we draw conclusions in Section 9.

2. RELATED WORK

2.1. Vision indoor robot localisation

The first vision systems developed for mobile robot localisation relied heavily on the geometry of space and other metrical information for driving the vision processes and performing self-localisation. In particular, interior space was represented by complete CAD models containing different degrees of detail. In some of the reported work [5], the CAD models were replaced by simpler models, such as occupancy maps, topological maps, or even sequences of images.

DeSouza and Kak [2] gather the existing approaches in three categories according to the apriori knowledge provided to the system:

(i) *map-based localisation:* these systems depend on user-created geometric models or topological maps of the environment;

(ii) *map-building-based localisation:* these are systems that use sensors to construct their own geometric or topological models of the environment and then use these models for localisation and navigation;

(iii) *mapless localisation*: these systems do not use any explicit representation of the environment. Rather, they are based on recognised objects found in the environment and the tracking of those objects by generating motions based on visual observations. Figure 3 resumes these categories and give maisn approaches within each one.

Most vision techniques for autonomous mobile robotics are map-based approaches, especially those based on absolute localisation which matches perceived data with an initial model to determine the robot position and those based on incremental localisation when the initial position of the robot is known. Incremental localisation methods use generally geometrical representation [6] or topological representation [7] of space. However, in large-scale and complex spaces, incremental localisation methods are not sufficiently accurate to determine the robot's position due to the accumulating error of the robot position's estimate. On the other hand, for absolute localisation methods, the step which establishes matches between robot's observation and features often stored in a geometrical-based model (expectation) is the most difficult among all steps in localisation systems and pose several problems. Moreover, if we consider a large-scale and complex space, matches between observation and expectation is increasingly difficult to solve. One can do localisation by landmark tracking when both the approximate location of the robot and the identity of the landmarks seen in the camera image are known and can be tracked. The landmarks used may either be artificial ones, such as stretched tapes and circles with a unique bar-code as reported by Tsumura in [8], or natural ones, such as doors, windows, and so forth. In this last case, this technique is related to object recognition methods.

Map-building-based systems allow robot to explore an unknown environment and build a map of that environment with simultaneous localisation and mapping (SLAM) methods. SLAM methods generate either topological [9] or geometrical representation of a space [10]. A challenging problem in map-building-based systems is the robot's ability to ascertain its location in a partially explored map or to determine that it has entered new territory. On the other hand, in mapless systems no maps are ever created. We usually call these systems as mapless navigation systems because of the needed robot motion purpose and the unknown absolute positions of elements of the environment. Indeed, relevant information about the elements in the environment are stored and associated with defined commands that will lead the robot navigation. Unlike this purpose, our global mapless localisation system aims rather to localise coarsely the robot and thus simplify the search space. It resembles appearance-based matching methods [11], but in our case we use image retrieval techniques to give a coarse estimate of the robot position. Thus, its outputs are one room label and one main orientation in this room. These characteristics make particular our approach (definition and results points of view).

2.2. Image retrieval systems

Content-based image retrieval (CBIR) systems have been essentially developed because the digitalised images databases are increasingly bulky. These images are, in general, compressed before being filed in databases. Once these data are stored, the problem is the capacity to retrieve them simply. An efficient reuse of these databases passes by the joint development of indexing and retrieving methods. A coarse representation of such a data management can be described as follows:

$$\{\text{image}\} \longrightarrow \text{features} \longrightarrow \text{indexing}. \tag{1}$$

The first systems suggested in the literature are based on the use of key words attached to images. The retrieving results of a particular type of image are inevitably a function of the lexical fields used. The indexing phase is, in this case, tedious and the coded data of the image remains limited. Thus, the content-based image retrieving is quickly developed giving rise to many systems allowing an image query method instead of the textual searching.

A content-based image retrieval system comprises generally four tasks. The principal ones are obviously the *indexing* and the *retrieving* tasks. The indexing task consists in computing a signature summarizing contents of an image which will be then used in the retrieving stage. The attributes usually used as signature are colour, texture, and shape. On the other hand, the retrieving task is generally based on a similarity measure between the signature of the request image and those in the corresponding database. We used only these two tasks for our automatic robot localisation problem. The two other tasks are *navigation* and *analysis*. Navigation is mainly related to the manner of database's consultation. This functionality is often static with a search for one or more answers

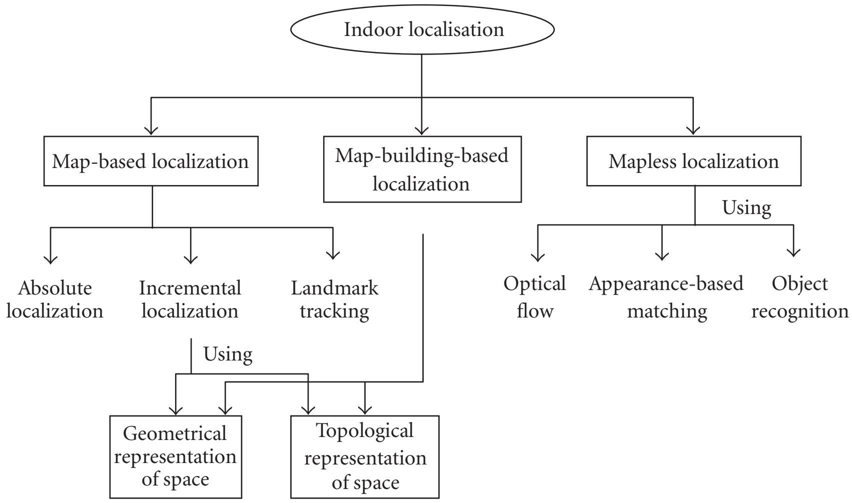


Figure 3: Robot localisation categories.

to a given request. A new type of research more interactively results in a more incremental approach and especially more adaptive to the users needs. From the retrieved images resulting from the first stage, the user can refine his research according to an object or a selected zone. This analysis is providing quantitative results and not of visual nature (e.g., the number of images with a blue colour bottom). This functionality is thus summarised to extract statistics from images.

In addition, image retrieval systems are generally based on a query by example (QBE): further to a request image taken by a robot in our case, the search engine retrieves the closest images of the database on the basis of a similarity distance. Then the ideal retrieving tool is that which quickly and simply gives access to the relevant images compared to a request image taken instantaneously by the mobile robot. The question is how to retrieve, automatically from the database, images visually similar to the request image. The similarity is evaluated by using a specific criterion based on colour, shape, texture, or a combination of these features. Many techniques were proposed with colour-based image retrieval [12–14], and it is impossible to define the best method without taking account of the environment. We can nevertheless release a general methodology through the following stages [15]:

(i) elicitation of a significant reference base allowing storing images and files of index associated with each image;
(ii) quantisation of each image by keeping only the relevant colours in order to optimise the efficiency in time and in results;
(iii) defining images signatures according to the desired requests (signature consists of a combination of generic attributes and specific attributes related to the application);
(iv) choice of a metric for the similarity measure;
(v) implementation of an interface allowing requests by image examples for the concerned applicability.

Many academic and/or industrial content-based image retrieval systems were developed: Mosaic [16], Qbic [17], Surfimage [18], Netra [19], VisualSEEK [20], and so forth. They allow an automatic image retrieving per visual similarity. The standard architecture of all these marketed systems comprises an offline phase to generate image's features and an online phase for image retrieving task (as showed by Figure 4).

Some systems are conceived for general public applications (e.g., the search of images on Internet). Image databases are then general and include heterogeneous type of images. Other systems are conceived for specific applications. The used image databases are in this case more concise and specific to the application. Images are characterised by homogeneous contents (faces, medical images, fingerprints, etc.). In the specific databases, the developed features are dedicated and optimal for the target considered (eccentricity of the contour of a face, position of a tumour, etc.). On the other hand, for the generic databases, the extracted features are universal (colour, texture, shape, etc.) [21]. Although our specific applicability (the global localisation of a robot in an indoor environment), image databases are generic because of the variety of objects present in a house and indoor spaces in general (see Figure 5).

3. IMAGE DATABASES

Two complete and well-structured image databases are built in two different indoor spaces (domestic environment) to assess the global localisation of the robot. Both spaces are large-scale and complex indoor environment owing to the fact that each of them contains 8 different rooms including the kitchen, the living room, and even the bathroom. Images of each database have been taken from all the rooms of the corresponding indoor space. For each room, we find a lot of images, corresponding to different available position of the robot and different orientation with a rotation of 20° or 30° according to the room dimensions. The first database contains 240 images and the second 586 images. The size of

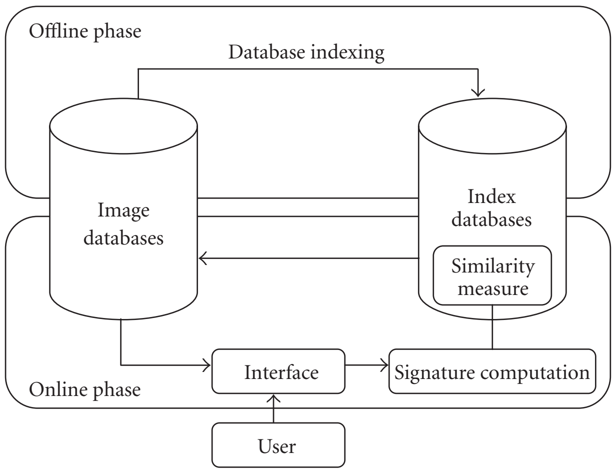


Figure 4: Content-based image retrieving architecture.

images is 960 × 1280 pixels. Figure 5 shows examples of images from the first database (a, b) and from the second one (c, d).

In the second database, we take also the luminosity into account (cf., Figures 5(c), 5(d)). For the same position, we have two or three images which have been taken at different day time. We also took a lot of request images which are different from the database images. For the first database, we have 20 request images and 35 for the second database.

4. COLOUR HISTOGRAMS

Colourimetric information is very significant in a domestic environment. Indeed, such a space includes various elements without colourimetric coherence between them. A discrimination of these elements can be more powerful by taking into account their colours.

Colour histograms remain the most used techniques as for adding colour information to retrieval systems. The robustness of this feature and its invariance to the position and orientation of objects make its strong points. Nevertheless, these performances are degraded quickly when the database is large. But in our application, the image database is not very bulky. Indeed, in an indoor environment, we do not exceed a few hundreds of images to describe structurally the environment of the robot. The use of the histograms for colour images indexing is based primarily on the selection's techniques of the adapted colour space, the quantisation of the selected space, and the comparison methods by similarity measures. We have tested the RGB and the LUV colour spaces. To the RGB colour space which gave best results, we developed several uniform quantisations in order to test different pallet sizes.

Given a colour image I, of size M by N pixels, the colour distribution of a colour bin $\mathbf{c}$ which ranges over all bins of the colour space is given by

$$h_c^I = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\delta(I(i,j)-c). \tag{2}$$

In the above equation, $\delta()$ is the unitary impulse function. We notice that the h_c values are normalised in order to sum to one. The value of each bin is thus the probability that the colour $\mathbf{c}$ appears in a pixel of the image. Different similarity measures were implemented and tested to our image databases. Two category of measures are presented: the bin-by-bin similarity measures which compare contents of corresponding histogram bins (Minkowski distance, histogram intersection, and the χ^2 test) and the cross-bin measures which compare noncorresponding bins (Mahalanobis distance and EMD Distance). Hereafter we present those similarity measures between a request image (I) and all the database images (H).

(1) *Minkowski distance:*

$$d(I,H) = \left(\sum_c |h_c^I - h_c^H|^r\right)^{1/r} \quad r \geq 1 \tag{3}$$

 (a) Manhattan distance L_1: $r = 1$
 (b) Euclidean distance L_2: $r = 2$

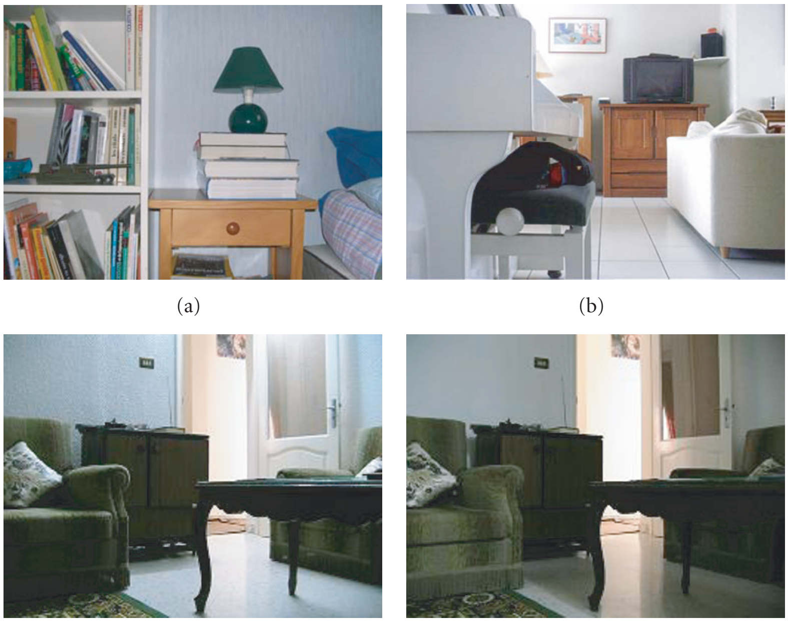
(2) *Histogram intersection:*

$$\text{Inters}(I,H) = \frac{\sum_c \min(h_c^I, h_c^H)}{\sum_c h_c^H}. \tag{4}$$

This function deducts the number of pixels of the model which have a direct correspondent in the request image. Values close to 1 indicate a good resemblance [12].

(3) *The χ^2 test.* A colour histogram can be considered as the realisation of a random variable giving colours in an image. Thus, the histogram comparison can be brought back to a test of assumptions, on which it is necessary to determine if two achievements (i.e., two histograms) can come from the same distribution. The χ^2 test is based on the assumption that the present distribution is Gaussian [22]. The χ^2 test is given by

$$\chi^2 = \sum_c \frac{(h_c^I - h_c^H)^2}{(h_c^I + h_c^H)^2}. \tag{5}$$



(a) (b) (c) (d)

Figure 5: Examples of indoor images.

(4) *Mahalanobis distance* or generalised quadratic distance D_{QG} was used by Niblack et al. [23] to take into account the intercorrelation between colour components. A weighting matrix W which includes the resemblance between colours was proposed. The generalised quadratic distance resulting from the Euclidean distance is defined by the following formula:

$$d_{QG}(I,H) = \sqrt{(H-I)W(H-I)^T}. \tag{6}$$

The components w_{ij} of the weighting matrix W can be interpreted like similarity indices between the i^e and *the* j^e element of the pallet. Thus W is generally represented by the reverse of the intercorrelation matrix between colour bins. Other proposals of weightings matrices attached to the representation of colour spaces were introduced by Striker and Orengo to define the colourimetric distances between colours [24].

(5) *EMD distance. Earth mover distance* proposed by Rubner et al. [25] consists in the extraction of the minimal quantity of energy necessary to transform a signature into another. Having the distances d_{ij} between colours components of the two histograms H and I of m and n dimensions, respectively, it is a question of finding a whole flow $F = [f_{ij}]$ which minimises the cost of the following quantity:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}. \tag{7}$$

To control the implied energy exchanges, the direction of transfer must be single ($f_{ij} \geq 0$) and a maximum quantity of transferable and admissible energy of each colour component should be defined. From the whole of optimal transfer F, EMD distance is then defined as the following resulting work:

$$d_{\text{EMD}}(H,I) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}. \tag{8}$$

The formalism suggested by Rubner meets all conditions to determine the optimal distance between two histograms but the complexity introduced by the algorithm of optimisation makes it complex in time computing [26].

5. A NEW COLOUR FEATURE DEFINITION

5.1. Baker's transformation

The baker's transform (BT for short) is based on the definition of mixing dynamical systems [27, 28]. The main interest of these transformations is that they mix in a very homogeneous way all the elements of the involved space.

Arnold and Avez [27] give a lot of examples of such mixing transformations, which are defined on the unit square $[0,1] \times [0,1]$. We have used one of them, the BT. We just mention here that all the examples given by Arnold and Avez are defined on continuous sets. On the other hand, digital images are finite sets of points (pixels). Unfortunately, it appears that a transformation of a finite set is never a mixing one. But for some peculiar mixing transformations like BT, even restricted to finite sets, pixels are statistically well mixed by a suitable number of iterations.

Figure 6: 256 × 256 original image.

Figure 7: First step of BT initial iteration.

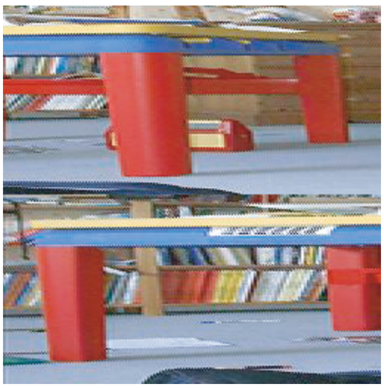


Figure 8: Second step of BT initial iteration.

An iteration of the BT is based on two steps:

(i) first, an "affine" transformation is used which gives an image twice larger and half higher (cf. Figure 7) from an original image (cf. Figure 6);

(ii) then, the resulting image is cut vertically in the middle and the right half is put on the left half (cf. Figure 8).

After a suitable number of iterations, we obtain a well-mixed image (cf. Figure 9). From this mixed image, we extract a definite size window (16 × 16 in the example) which gives after some iterations a reduced scale version of the original image (cf. Figure 10). The BT requires that the image size is $2^N \times 2^N$ pixels and we can show that the BT is periodic with period equal to $4N$ iterations. The image is well mixed with N iterations. If we divide the mixed image and take a $2^p \times 2^p$ resulting window ($P < N$), we can obtain a good version of the original image at a reduced scale after applying $3p$ iterations of the BT to the mixed $2^p \times 2^p$ window.

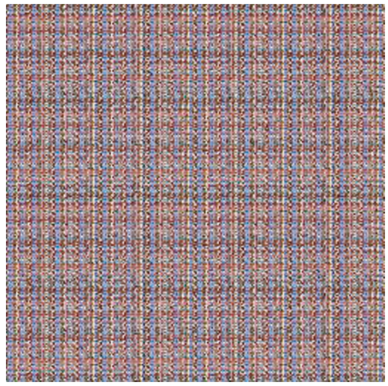


Figure 9: Well-mixed image.

Figure 10: 16 × 16 pallet deduced from the mixed window.

5.2. *The colour feature*

As shown in Figure 10, a small image of size 16 × 16 gives a good colour, shape, and texture representation of the original image and we can consider it as a representative colour pallet. In [29], we presented a first use of this method to quantify colour images. The idea is to use one of these windows as a colour pallet to reduce all the colour levels of the original image. With a $2^N \times 2^N$ image, it is possible to propose pallets containing 2^{2p} colours ($P < N$). So the number of different pallets available from one image is given by the number $K = 2^{2(N-p)}$. Given a pallet, the common principle is, for each pixel, to compute the Euclidean distance between its colour and all colours present in the pallet. Then the new colour assigned to the pixel is that which minimises the distance. The problem is how to choose the representative window to build the good pallet? We analyse four different solutions and we show that the best of them uses selection of "the median pallet." The evaluation of results is done by a similarity distance between the original image and the reduced one. This distance, baptised "delta," is computed on each of the three colour channels (red, green, and blue) for all image pixels; in (9), I_1 and I_2 represent, respectively, the colour levels of a pixel in the initial image and in the reduced image:

$$\text{delta} = \frac{\sum_{i=1}^{2^N} \sum_{j=1}^{2^N} \left| I_1(i,j) - I_2(i,j) \right|}{2^N \times 2^N}. \tag{9}$$

From a practical point of view, BT is a space transformation. For a given dimension of image, the position of the output pixels in the mixed image is always the same one.

Table 1: "delta" distance between request image and reduced ones.

Figure	delta R	delta V	delta B	<delta>
8(a)	4.01	4.12	5.19	4.44
8(b)	73.19	30.49	23.86	42.52

Table 2: Results for database n°1–20 request images.

	Colour number	48	108	192	300	588	%
First answer	*Right*	5	9	8	9	9	**40**
	Medium	6	3	4	4	2	**19**
	False	9	8	8	7	9	**41**
Three answers	*Right*	10	11	13	13	13	**20**
	Medium	24	21	17	18	21	**33.7**
	False	26	28	30	29	26	**46.3**

Consequently, a look up table (LUT), which indicates for each pixel of an image its coordinates in the mixed image, allows to obtain the pallet more quickly. In another way, BT simply consists to extract in a homogeneous way pixels from the image. Thus, it is possible, for rectangular images, to obtain a same feature by applying a subsampling technique.

6. RETRIEVAL APPROACHES

6.1. Colour reduction retrieval approach

If it is possible to extract a sample of pixels, which the colours are representative of the original image and which are stable for images having the same sight, then this feature is called colour invariant. This colour feature is used as an indirect signature [30]. The strategy to retrieve the closest image from the database, to the request image, is shown in Figure 11. First we build a pallet database by computing for each image of the original database its colour invariant. Then, the request image is projected in the colour space defined by each pallet from this pallet database. We compute the colour difference between the request image and the projected ones (cf. Table 1), and we select the pallet (i.e., the image) which leads to the minimum of this distance.

6.1.1. Results of the colour reduction retrieval approach

From each image database, we have built 5 pallet databases, to assess different size of pallet: 48, 108, 192, 300, and 588, which, respectively, correspond to these two dimensional pallets of: 6×8, 9×12, 12×16, 15×20, and 21×28. In order to speed up the retrieval process, we subsampled the request image (60×80 pixels). Tables 2 and 3 display a synthesis of obtained results. The retrieved images are organised in three classes.

(i) Right: the image proposed by the retrieval system is taken in the same room and with the same orientation than the request image.

Table 3: Results for database n°2–35 request images.

	Colour number	48	108	192	300	588	%
First answer	*Right*	10	16	17	21	19	**47.5**
	Medium	13	7	12	6	7	**25.7**
	False	12	12	6	8	9	**26.8**
Three answers	*Right*	23	35	37	37	35	**31.8**
	Medium	43	32	36	37	38	**35.4**
	False	39	38	32	31	32	**32.8**

(ii) Medium: the image proposed by the retrieval system is taken in the same room than the request image.

(iii) False: the image proposed by the retrieval system is taken in other room than the request image.

We analysed two cases: the quality of the first answer and the quality of the three first answers. We can see that we obtain 40% or more of good answers when we take only one answer into account. If we want a coarse answer to the question "In which room is the robot"?, we sum the "Right" and the "Medium" answers. Then the rate of correct answer is about 60% for the database n°1 and over 70% for the second database. When we take the first three answers into account, we obtain degraded results especially for the first database which contains no more than one image for each sight.

Moreover, the relationship between accuracy and colour number is not monotonic. Above a certain threshold, performance gains from increased colour number cease to be observed and become too small to justify the increased computational cost. In the second database, we obtain results over 75% with 192 and 300 colours in the pallet. Finally, we retain this last size (300 colours) to work with for the next experiments.

Figures 12(a) and 13(a) show request images from the first and the second databases, respectively. Figures 12(b), 12(c), and 12(d) present the first three answers obtained (Figures 12(b) gives the right response, Figures 12(c) and 12(d) are false). Figures 13(b) and 13(c) present two examples of the first answer obtained with two different pallets. We can see that the result is right with a pallet of 192 colours (see Figure 13(b)), but it is false with a pallet of 48 colours (see Figure 13(c)).

In spite of its interest which validates the concept of colour invariant, our method is handicapped by a very significant computing time (over than 15 minutes). The projection of the request image according to all pallets of the database takes a more and more time that the bulky database. We can however consider the pallet as a feature and compare pallets between them in the retrieving phase instead of comparing request image with reduced ones.

6.2. The interpallet distance

After a first use of this colour pallet as an indirect descriptor, we associate to this feature an Euclidean distance that we call interpallet distance L_2 $(P_{\text{req}} - P_{\text{base}})$ [31]. The strategy to

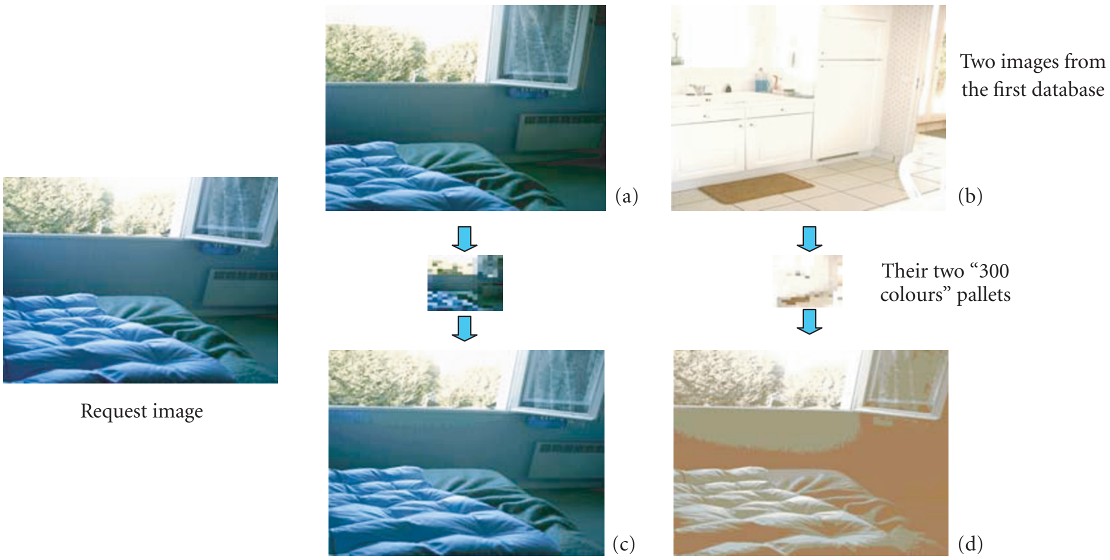
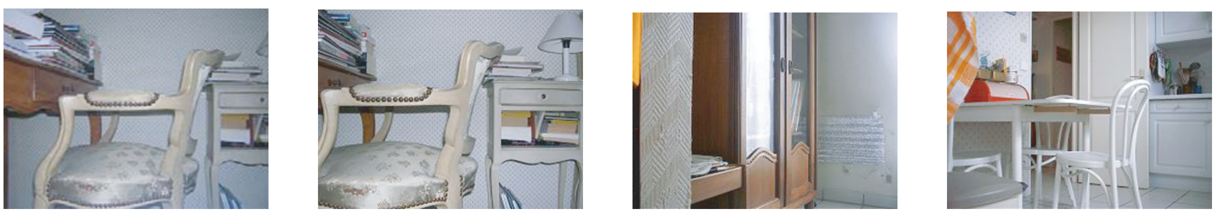


FIGURE 11: Request image reduced by pallets of the images (a) and (b) give the result images (c) and (d), respectively.



(a) (b) (c) (d)

FIGURE 12: Three answers with a pallet of 300 colours from the request image (a).

search the closest image to the request image is described as follows (cf. Figure 14).

(i) First we build a pallet database by the computation of the colour invariant of each image from the original database.

(ii) Then, we extract the pallet of the request image to compute the colour difference between this one and all pallets already built in the database. Euclidean distance is computed between correspondent colour having the same position in these pallets.

(iii) Finally, we select the pallet (i.e., the image) which leads to the minimum of this distance.

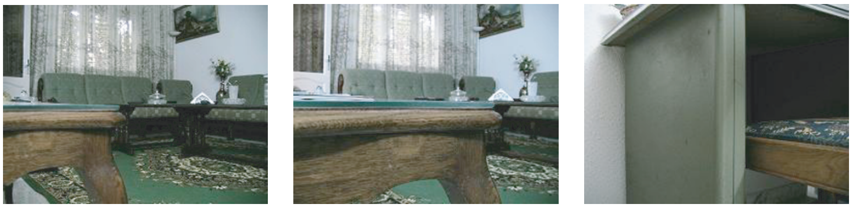
The space organisation of colours of this two-dimensional pallet is an additional information who can present invariance property to some changes in image sample. Thus, we emphasis this colour feature aspect and try to model it by preserving the interpallet distance which gives interesting results. Indeed, as the below figure shows it, the pallet preserves the spatial distribution and the principal vicinity relations between colours present in the original image. This should give us a relative invariance as well for sight point small changes as for scale factor (i.e., distance separating the camera to objects).

6.3. Space distribution of colours

In order to coarsely describe colours distribution form of the image and to build an invariant feature as well for sight point small changes as for scale factor, we extract the three first colour statistical moments of the pallet. These moments are largely used in pattern recognition systems and give a robust and complete description of analysed patterns. Stricker and Orengo [24] establishes a balanced sum of the average, the variance, and skewness (the third-order moment) computed for each colour channel, to provide a single number used in the indexing process. These moments are defined by

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij},$$
$$\sigma_i = \frac{1}{N}\sqrt{\sum_{j=1}^{N}\left(p_{ij}-\mu_i\right)^2}, \tag{10}$$
$$s_i = \frac{1}{N}\left(\sum_{j=1}^{N}\left(p_{ij}-\mu_i\right)^3\right)^{1/3},$$

where p_{ij} is the value of the pixel j in the colour channel I, N is the number of pixel in the image.



(a) (b) (c)

Figure 13: First answer with a pallet of 192 colours (b) and 48 colours (c) from the request image (a).

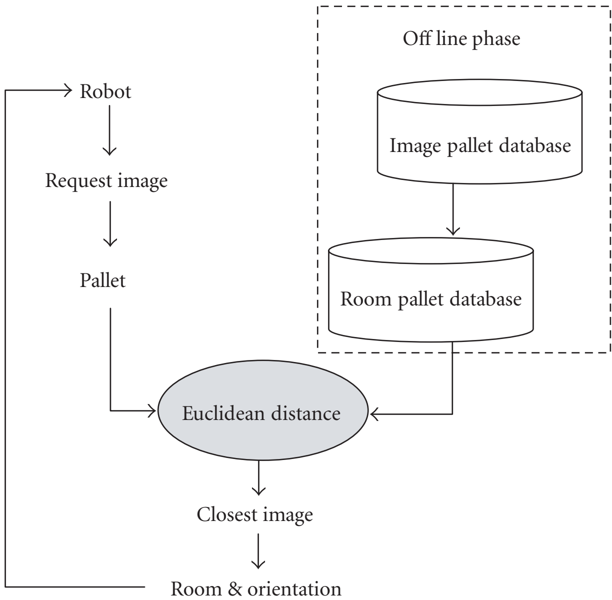


Figure 14: Interpallet distance.

The distance between two images is then defined like a weighted sum between these quantities for each channel:

$$d_{\text{mom}}(I,H) = \sum_{i=1}^{3} w_{i1}\left|\mu_i^I - \mu_i^H\right| + w_{i2}\left|\sigma_i^I - \sigma_i^H\right| + w_{i3}\left|s_i^I - s_i^H\right|. \tag{11}$$

We have applied these moments on our two-dimensional pallet. p_{ij} are in this case pixels from the pallet and N is the number of colour in the pallet. We notice that a space description of our two-dimensional pallet by colour moments as showed in [20], gives better results than a similar description of the entire original image. We deduce that such a description of a pallet, which is a represention on a reduced scale of the original image, gives a more precise visual summary of it. In addition, the search time is much more faster while operating on pallets (0,7 second against 3 to 4 seconds for retrieving by image moments with an image size of 1260×960 pixels).

Nevertheless, the success rate remains rather weak compared to our objectives (50% to find the right room). Thus, we studied the discriminating capacity of each of the first four moments (average, variance, skewness, and kurtosis) to use the best of them as a weighting factor to the proposed interpallet distance. After the computation, the first four moments variance, the greatest on is used to build a weighting coefficient enough discriminating for strong variations and neutral for weak variations (lower than a threshold α). Then we discriminate through the coefficient λ images having a variance of the first two moments lower than a threshold β. Following some experiments on our two image databases, we fixed α at 20 and β at 128:

$$w_1 = \lambda \frac{\Delta\sigma}{\sigma_{\text{im}} + \sigma_{\text{req}}} \tag{12}$$

with

$$\Delta\sigma = \begin{cases} \alpha & \text{if } \left|\sigma_{\text{req}} - \sigma_{\text{im}}\right| < \alpha, \\ \left|\sigma_{\text{req}} - \sigma_{\text{im}}\right| & \text{otherwise,} \end{cases} \tag{13}$$

$$\lambda = \begin{cases} 1 & \text{if } \left|\sigma_{\text{req}} - \sigma_{\text{im}}\right| < \beta,\ \left|\mu_{\text{req}} - \mu_{\text{im}}\right| < \beta, \\ \infty & \text{otherwise.} \end{cases} \tag{14}$$

Thus

$$D_1 = w_1 \cdot L_2\left(P_{\text{req}} - P_{\text{im}}\right). \tag{15}$$

6.4. *Vicinity template of colours*

To describe the textural aspect of colours distribution, we developed the cooccurence matrix and some relating features defined by Haralick et al. [32] and extended to colour information by Trémeau [33] which are

(i) colour inertia:

$$I = \sum_{i=0}^{N}\sum_{j=0}^{N} D_{ij}^2 \cdot P(i,j) \tag{16}$$

with $D_{ij}^2 = (R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2$; R, G, and B are the three colour channels of the RGB colour space;

(ii) colour correlation:

$$C = \sum_{i=0}^{N}\sum_{j=0}^{N} \frac{D_i \cdot D_j}{\sigma_i \cdot \sigma_j} P(i,j) \tag{17}$$

with $D_i = \left((R_i - R_{\mu_i})^2 + (G_i - G_{\mu_i})^2 + (B_i - B_{\mu_i})^2\right)^{1/2}$, $D_j = \left((R_j - R_{\mu_j})^2 + (G_j - G_{\mu_j})^2 + (B_j - B_{\mu_j})^2\right)^{1/2}$ with

μ_i, σ_i (resp., μ_j, σ_j) who represent the colour average and the colour standard deviation for all the transitions for which the index colour first pixel is i (resp., the index colour second pixel is j).

Thus $\mu_i = (R_{\mu_i}, G_{\mu_i}, B_{\mu_i})$ with

$$R_{\mu_i} = \frac{1}{\sum_{j=0}^{N} P(i,j)} \cdot \sum_{j=0}^{N} P(i,j) \cdot R_j,$$
$$\sigma_i = \sqrt{\frac{1}{\sum_{j=0}^{N} P(i,j)} \cdot \sum_{j=0}^{N} P(i,j) \cdot D_j^2} \tag{18}$$

(iii) homogeneity:

$$H = \sum_{i=0}^{N} \sum_{j=0}^{N} P(i,j)^2 \tag{19}$$

(iv) entropy:

$$E = \sum_{i=0}^{N} \sum_{j=0}^{N} P(i,j) \cdot \log P(i,j). \tag{20}$$

Moreover, we extract the maximum value of the cooccurence matrix and its two colour components that we note (c_1, c_2).

Owing to the fact that a fine quantisation of a colour space gives a large signature, the construction of cooccurence matrices related to pallets (low-size images) brings smooth and not enough discriminating distributions. To mitigate this problem, we kept only the main colours vicinity and we developed a new cooccurence matrix related to a coarse uniform quantisation of the RGB colour space in 64 bins. We considered, in addition, an isotropic vicinity (8 connexities) of each pixel.

For the retrieval phase, we developed the Euclidean distance $L_2(M_{\text{req}} - M_{\text{im}})$ between cooccurence matrices M_{req} and M_{im}, respectively, of request image pallet and database image pallet. This distance is weighted with the factor w_2 computed on the base of the entropy variation which has the greatest dynamics and so the discriminating capacity among the other cooccurrence matrix features. This gives the D_2 distance hereafter:

$$w_2 = \lambda \frac{\Delta E}{E_{\text{im}} + E_{\text{req}}} \tag{21}$$

with

$$\Delta E = \begin{cases} \gamma & \text{if } |E_{\text{req}} - E_{\text{im}}| < \gamma, \\ |E_{\text{req}} - E_{\text{im}}| & \text{otherwise.} \end{cases} \tag{22}$$

Thus

$$D_2 = w_2 L_2(M_{\text{req}} - M_{\text{im}}). \tag{23}$$

We analysed the colour components of the maximum value of the cooccurence matrix. We estimated the value λ according to the three-dimensional connexity of the request image colour components and those of database images. By assimilating uniform colour bins to cubes (cf. Figure 15), the three-dimentional connexity is evaluated as follows:

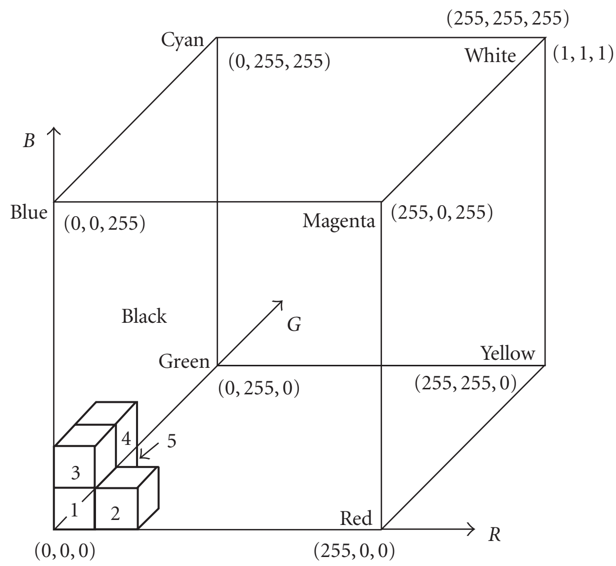


Figure 15: RGB colour cube.

Connex1: cubes adjacent by a surface, for example, cubes $(1, 2)$, $(1, 3)$;

Connex2: cubes adjacent by an edge, for example, cubes $(2, 3)$, $(3, 5)$;

Connex3: cubes adjacent by a point, for example, cubes $(2, 4)$.

We retained the best connexity of the request image pair of colour $(c_1, c_2)_{\text{req}}$ with each database image pair of colour $(c_1, c_2)_{\text{im}}$ through the following algorithm:

if $(c_1, c_2)_{\text{req}} = (c_1, c_2)_{\text{im}}$, then $\lambda = 1$,
if $(c_1, c_2)_{\text{req}}$ Connex1 $(c_1, c_2)_{\text{im}}$, then $\lambda = 2$,
if $(c_1, c_2)_{\text{req}}$ Connex2 $(c_1, c_2)_{\text{im}}$, then $\lambda = 3$,
if $(c_1, c_2)_{\text{req}}$ Connex3 $(c_1, c_2)_{\text{im}}$, then $\lambda = 4$,
else $\lambda = 8$.

6.5. *The final distance* D

The D_2 distance built by cooccurence matrices of the pallets gives lower results than the D_1 distance (only 55% of right room). But by finely analysing answers of each request, we note that there are some cases where the D_1 distance led to a false result whereas the distance D_2 leads to a right result (and vice versa).

The final distance D we propose takes the normalised distances between pallets and between cooccurence matrices into account, each one balanced by a resulting term from colour moments and cooccurence matrices attributes, respectively. The balanced sum distance D is given by

$$D = w_1 L_2(P_{\text{req}} - P_{\text{im}}) + w_2 L_2(M_{\text{req}} - M_{\text{im}}). \tag{24}$$

6.6. *Hierarchical approach*

We proposed as preliminary stage, before applying the proposed distance D, a hierarchical search using classification of images according to rooms. We characterise each room by

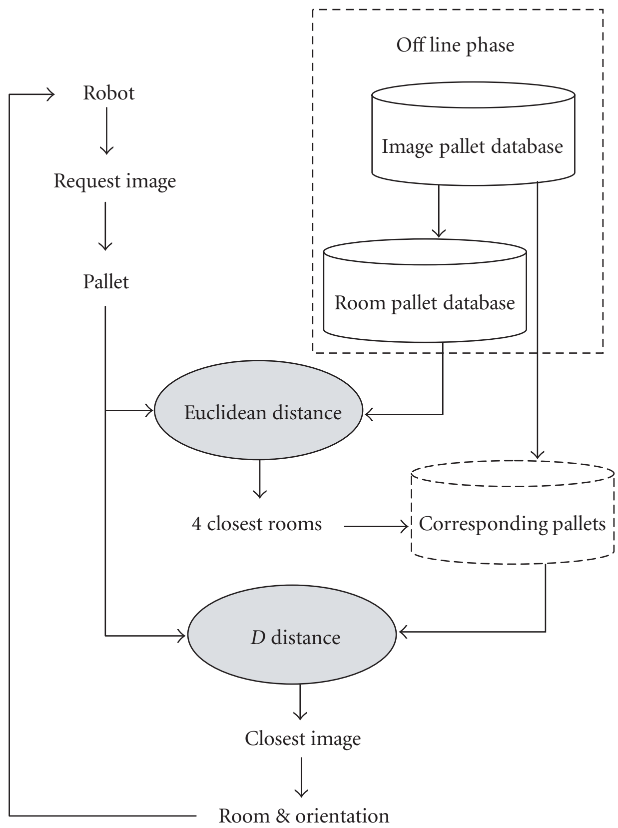


Figure 16: Hierarchical search.

a discriminating colour pallet. Each room pallet is built by sampling colour pallets of images belonging to this room and by adding colours whose minimal distance to those retained is higher than a threshold fixed at 10.

During the retrieval phase, we compute the difference between the request and room pallets by using the Euclidean distance and then classify these distances by ascending order. We eliminate firstly rooms presenting no similar colours to those of the request image (cf. Figure 16). Finally, we apply the distance D to retrieved images where the robot is high probably lost.

To increase the speed of the system it is necessary to eliminate the maximum of rooms. However, we should not affect the system effectiveness by eliminating the room corresponding to the request image where the robot is lost. After some experiments on our two image databases, we kept in search process the first four rooms given by the hierarchical process. It should be noted that our aim is to simplify the large-scale indoor space to a simple part (some rooms) where a fine localisation system should be more efficient. This being made, we eliminate, from the localisation process, rooms on which certain images distort results. By keeping the first four rooms from eight rooms in the indoor environment, we divide by two the space where one seeks to locate the robot.

Usual localisation contributions assess their system in simple indoor environments with a maximum of two rooms to draw results and conclusions of a robot's exact localisation. We evaluate our method in two indoor environments

(a) (b)

Figure 17: (a) Request image from the second database; (b) Response image within D distance.

much more large and complex. We do not propose a solution of exact localisation, but rather a solution to simplify the complexity of the space. If we want to make robot's fine localisation with a map-based method, for example, our algorithm can simplify from the search half of the map without compromising any result. Indeed, for our both assessment databases, we can undoubtedly simplify 4 of the 8 rooms of the indoor space without eliminating the required room. These results make of this hierarchical process an effective preliminary stage not only for our localisation method but for all indoor localisation systems. Moreover, all the process does not take more than two seconds to have half of the indoor space simplified.

7. EXPERIMENTAL RESULTS

We performed various experiments with 20 test images for the first database and 35 for the second one. Test images are different from those of the data sets. We note that a system which guesses the right room and orientation of the robot would be right one out of hundred times, giving an error rate of 99%.

We note better results with our distance D than the interpallet distance owing to the fact that we consider as well the space organisation and colour vicinity as the colourimetric aspect of the pallet. For the request image (cf. Figure 17(a)), we have false results by using the interpallet, D_1 and D_2 distance separately. The distance D combining these two last measures gives the right result. We have in this case the result image of Figure 17(b) indicating the right room and the right orientation.

Table 4 shows results of our retrieval strategies based on a two-dimensional colour pallet extract from the Baker's transformation, showing the advantages of our approach.

The first global retrieving approach based on the colour reduction principle, while giving acceptable results (about 70%) proved limited by a prohibitory computing time. We developed, thereafter, a method by our colour pallet description with an interpallet distance. Results were a little degraded while reducing the computing time appreciably. Seeking to optimise the quality of description as well as the computing time, we took into account the space organisation of the pallet to define a specific new similarity measure. Thus we could improve the results with a search time of about three seconds.

Table 4: Results of our methods.

	Database 1				Database 2				Databases1 & 2
	Time (s)	Results	%		Time (s)	Results	%		%
Colour reduction retrieval approach									
Right	0.7	9	45%	65%	1.1	21	60%	77.2%	71.5%
Medium		4	20%			6	17.2%		
False		7		35%		8		22.8%	28.5%
Interpallet distance									
Right	0.7	10	50%	60%	1.1	15	42.8%	60%	60%
Medium		2	10%			6	17.2%		
False		8		40%		14		40%	40%
Spatial and colour distance D									
Right	2	12	60%	65%	3	18	51.5%	71.5%	69%
Medium		1	5%			7	20%		
False		7		35%		10		28.5%	31%
Hierarchical approach with the distance D									
Right	4	12	60%	70%	5	20	57%	88.5%	82%
Medium		2	10%			11	31.5%		
False		6		30%		4		11.5%	18%

In order to improve even more the performances of room identification, we developed a hierarchical retrieving method eliminating in a preliminary stage a number of rooms from the indoor environment to combine speed and effectiveness of the localisation process. Results are clearly improved to exceed the 80%. The hierarchical procedure being consuming in computational time, the computing time of the global solution tends to increase to reach 4 seconds, a time considered to be acceptable for the task of global localisation.

In order to validate this work, we compare these results with a classical image retrieval technique which uses colour histogram. We developed colour histograms on RGB and Luv spaces. The RGB colour space which gives best results is performed to three uniform quantisations into 64, 512, and 4096 colour bins. The various bin-by-bin (histogram intersection, L_2 and χ^2) and cross-bin (earth mover distance and Mahalanobis distance) similarity measures developed previously were implemented and tested to our image databases.

As showed in Figure 18, the quantisation to 64 colours proves very coarse for colour histograms. Quantisation to 512 bins improve considerably these results but the 4096 bins discretisation gives the best results except the Euclidean distance which gives best results with 512 bin quantisation. We display results of the 4096 colour histograms in Table 5. We note the worst results with the cross-bin similarity measures which tend to overestimate the mutual similarity of colour distributions. Moreover, the computational complexity of the EMD and Mahalanobis distance are the highest among the evaluated measures. Indeed, computing the EMD between 4096 colour histograms in our database takes over than 30 minutes. The χ^2 test gives the best results among the five developed distance. This statistical measure gives an error rate of 22% to find the right room. In addition, computing time at around 4 seconds is acceptable for a global localisation task.

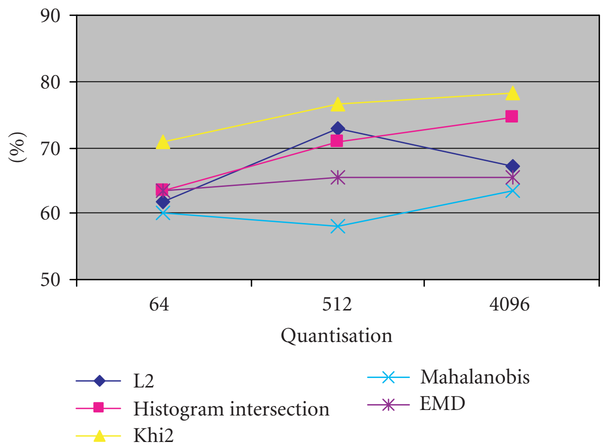


Figure 18: Histogram results.

Our method gives better results than those of the colour histograms. We have especially best results than the effective χ^2 test. For the second database which integrates different illumination conditions, we have a rate of 88% to find the right room giving a correct estimate of the robot position. Results provided only with a colour-based description of indoor images are encouraging. A final system obviously must integrate other type of signature like shape and texture with a more structured model of the environment.

8. FURTHER RESEARCH

We can identify the following avenues for improving performances.

(1) A first prospect for image retrieving is to develop a local searching approach (to images). A combination

Table 5: Histogram results.

	Base 1 & 2			
	Time (s)	Results		
Histogram Intersection				
Right	4	31	56.4%	74.5%
Medium		10	18.1%	
False		14		25.5%
Euclidean Distance L_2				
Right	4	27	49.1%	67.2%
Medium		10	18.1%	
False		18		32.8%
χ^2 test				
Right	4	32	58.2%	78.2%
Medium		11	20%	
False		12		21.8%
Mahalanobis Distance				
Right	60	20	36.4%	63.6%
Medium		15	27.2%	
False		20		36.4%
Earth Mover Distance: EMD				
Right	2100	23	41.8%	65.4%
Medium		13	23.6%	
False		19		34.6%

of the global solution (whole image processing) and the local approach may improve our system's performances. The characteristics developed in this paper were computed globally in the entire image. However, a system only based on global characteristics cannot give the desired results. Indeed, an image contains many objects having very different characteristics (colours and textures), the feature vector extracted from the whole image loses local information (related to objects) and gives a coarse idea about images' contents. A possible solution consists on indexing some known and nonremovable objects in the room's house [34]. A preliminary retrieving phase could determine whether one of these objects is in the sight field of the robot reducing the size of retrieving space. Such a combination of the global solution with a preliminary local approach has to improve even more the performances of our system.

(2) More careful modelling of the colour distribution for our similarity measure, for example, by using Tamura signature features [35] like directivity, contrast, and coarseness with a more fine colour pallet, or a frequency-based model [36] can introduce texture-useful information to improve discrimination between images.

(3) Another prospect for image retrieving problematic would consist on the exploration and the search for other features and invariants such as differential invariants for colour images and invariants for predictable change of illumination [37]. A comparison between our results and those gotten by these approaches could induce ideas to improve results.

(4) We could consider a procedure of reinforcement of the decision-making by asking the robot to take a second image of its environment (after a small translation and rotation) and by comparing the response obtained to that resulting from the first request image. A confidence factor attached to the answer could achieve this procedure effectively. It will be necessary, in any event, to seek a compromise between the quality of the results and the response time.

9. CONCLUSION

We have presented a new approach by image retrieval techniques which aims to localise an indoor robot. This approach uses a pallet extracted by using baker's transformation. This pallet gives a good representation of initial colours and preserves the spatial organisation of the original image. We also build an appropriate distance which integrates the space and the colour aspects of this pallet in order to find the closest image. We obtain results which are better than results obtained from a colour histogram method. Thus we have developed one retrieval technique which is fast and effective.

REFERENCES

[1] J.-J. Gonzalez-Barbosa and S. Lacroix, "Localisation d'un robot mobile dans un environnement naturel par indexation d'images panoramiques," in *LAAS/Robotique et Intelligence Artificielle Seminars*, Toulouse, France, Decemebr 2001.

[2] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[3] O. Ait-Aider, *Localisation référencée modèle d'un robot mobile d'intérieur*, Ph.D. thesis, Université d'Evry, Val d'Essonne, France, 2002.

[4] P. Hoppenot, E. Colle, O. Ait-Aider, and Y. Rybarczyk, "ARPH: assistant robot for handicapped people—a pluridisciplinary project," in *Proceedings of the 10th IEEE International Workshop on Robot and Human Communication*, pp. 624–629, Bordeaux, Paris, France, September 2001.

[5] R. Chatila and J.-P. Laumond, "Position referencing and consistent world modeling for mobile robots," in *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 138–145, St. Louis, Mo, USA, March 1985.

[6] A. Kosaka and A. C. Kak, "Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 56, no. 3, pp. 271–329, 1992.

[7] M. Meng and A. C. Kak, "Mobile robot navigation using neural networks and nonmetrical environment models," *IEEE Control Systems Magazine*, vol. 13, no. 5, pp. 30–39, 1993.

[8] T. Tsumura, "Survey of automated guided vehicle in a Japanese factory," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 3, pp. 1329–1334, San Francisco, Calif, USA, April 1986.

[9] H. Choset and K. Nagatani, "Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 125–137, 2001.

[10] H. P. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 2, pp. 116–121, St. Louis, Mo, USA, March 1985.

[11] S. D. Jones, C. Andresen, and J. L. Crowley, "Appearance based processes for visual navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '97)*, vol. 2, pp. 551–557, Grenoble, France, September 1997.

[12] M. Swain and D. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[13] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.

[14] J. R. Smith and S.-F. Chang, "Local color and texture extraction and spatial query," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '96)*, vol. 3, pp. 1011–1014, Lausanne, Switzerland, September 1996.

[15] C. Fernandez-Maloigne, "Quantification et segmentation pour l'indexation de bases d'images couleur," in *Traitement et Analyse d'Images: Méthodes et Applications (TAIMA '01)*, pp. 61–66, Hammamet, Tunisia, October 2001.

[16] S.-T. Goh and K.-L. Tan, "MOSAIC: a fast multi-feature image retrieval system," *Data & Knowledge Engineering*, vol. 33, no. 3, pp. 219–239, 2000.

[17] M. Flickner, H. Sawhney, W. Niblack, et al., "Query by image and video content: the QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[18] C. Nastar, "Indexation et recherche d'images: enjeux, méthodes et perspectives," Congrès IDT, 1999.

[19] W.-Y. Ma and B. S. Manjunath, "NeTra: a toolbox for navigating large image databases," *Multimedia Systems*, vol. 7, no. 3, pp. 184–198, 1999.

[20] J. R. Smith and S.-F. Chang, "VisualSEEk: a fully automated content-based image query system," in *Proceedings of the 4th ACM International Multimedia Conference*, pp. 87–98, Boston, Mass, USA, November 1996.

[21] N. Boujemaa, S. Boughorbel, and V. Constantin, "Description de la répartition spatiale de la couleur pour l'indexation d'images," in *13eme Congrès Francophone ARIF-AFIA (RFIA '02)*, vol. 2, pp. 405–414, Angers, France, January 2002.

[22] P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid, and G. Mistler, "Utilisation de la couleur pour l'appariement et l'indexation d'images," Research Report 3269, topic 3, Institut National de Recherche en Informatique et en Automatique (INRIA), Le Chesnay, France, 1997.

[23] W. Niblack, R. Barber, E. Equitz, et al., "QBIC project: querying images by content, using color, texture, and shape," in *Storage and Retrieval for Image and Video Databases*, vol. 1908 of *Proceedings of SPIE*, pp. 173–187, San Jose, Calif, USA, February 1993.

[24] M. A. Stricker and M. Orengo, "Similarity of color images," in *Storage and Retrieval for Image and Video Databases III*, vol. 2420 of *Proceedings of SPIE*, pp. 381–392, San Jose, Calif, USA, February 1995.

[25] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proceedings of the 6th IEEE International Conference on Computer Vision*, pp. 59–66, Bombay, India, January 1998.

[26] A. Trémeau, C. Fernandez-Maloigne, and P. Bonton, *Image Numérique Couleur: De L'acquisistion au Traitement*, Dunod, Paris, France, 2004.

[27] V. I. Arnold and A. Avez, *Problème Ergodiques de la Mécanique Classique*, Monographies Internationales de Mathématiques Modernes, Gauthier-Villars, Paris, France, 1967.

[28] P. Billingsley, *Ergodic Theory and Information*, John Wiley & Sons, New York, NY, USA, 1965.

[29] C. Montagne, S. Lelandais, A. Smolarz, and P. Cornu, "Adaptive color quantization using the "Baker's Transform"," in *Proceedings of the 2nd European Conference on Color in Graphics, Imaging, and Vision and Sixth International Symposium on Multispectral Color Science (CGIV '04)*, pp. 353–358, Aachen, Germany, April 2004.

[30] C. Montagne, S. Lelandais, A. Chaari, and M. B. Ahmed, "Invariant couleur basé sur la transformée du boulanger—application à la localisation globale d'un robot d'intérieur," in *International Conference Sciences of Electronics, Technology of Information and Telecommunication (SETIT '05)*, Susa, Tunisia, March 2005.

[31] S. Lelandais, A. Chaari, A. Smolarz, C. Montagne, and B. Jacquin, "A new color invariant for image retrieval using the baker's transformation," in *Beijing International Conference on Imaging: Technology and Applications for the 21st Century (BICI '05)*, pp. 328–329, Beijing, China, May 2005.

[32] R. M. Haralick, K. Shanmugan, and I. Dinsytein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[33] A. Trémeau and B. Gérin, "Utilisation de la matrice de cooccurrence couleur pour l'analyse de texture," in *Automatic Conference. Data-Processing Genius. Image (AGI '96)*, pp. 303–306, June 1996.

[34] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[35] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[36] K. S. Thyagarajan, T. Nguyen, and C. Persons, "A maximum likelihood approach to texture classification using wavelet transform," in *Proceedings of IEEE International Conference on Image Processing (ICIP '94)*, vol. 94, pp. 640–644, Austin, Tex, USA, November 1994.

[37] V. Gouet and P. Montesinos, "Normalisation des images en couleur face aux changements d'illumination," in *13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA '02)*, pp. 415–424, Angers, France, January 2002.