Global interior robot localization by a color content image retrieval system

Anis Chaari, Sylvie Lelandais, Christophe Montagne, Mohamed Ben Ahmed

Global interior robot localization by a color content image retrieval system

A. CHAARI[1,2], S. LELANDAIS[1], C. MONTAGNE[1], 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
Chaari.anis@ibisc.fr s.lelandais@iut.univ-evry.fr

Abstract
In this work we propose a new approach to determine the global position of a mobile robot in structured indoor area using content based image retrieval techniques based on color. We use, an original method of color quantization based on the baker's transformation to extract a two dimensional color pallet combining as well space and vicinity informations as colorimetric aspect of the original image. We conceive several approaches of research bringing to a specific similarity measure D integrating the space organization of colors in the pallet. The baker's transformation provides a quantization of the image into a space where colors 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 added to our system a hierarchical search module based on the logic classification of images following rooms allowing to eliminate some rooms from research and ensuring an improvement of our system performances. Results are then compared with those bringing by color histograms provided with several similarity measures. In this paper we focus on color based features to describe indoor images. A final system obviously must integrate other type of signature like shape and texture.

Key words: *content based image retrieval, quantization, baker's transformation, color pallet, space organization, co-occurence matrix, similarity measure, hierarchical search, color histogram.*

1. Introduction

The autonomous robot navigation in a structured interior or not structured external environment requires the integration of many functionalities, which go from the navigation control to the mission supervision, while passing by the perceived environment modelling and the planning of trajectories and strategies of displacements [1]. Among these various functionalities, the robot localization, i.e. the capacity to estimate constantly its position is very significant. Indeed, the knowledge of the robot position is essential to the correction of the trajectory and the execution of planned tasks.

Sensors constitute the fundamental elements of a localization system. According to the localization type 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 to estimate the current situation of the robot compared to its initial situation. On the other hand, the exteroceptive sensors measures the absolute situation of the robot by observing benchmarks whose situations is known in an environment frame attached reference.

The localization 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 not structured environments. These techniques can be gathered in two principal categories: relative localization methods and absolute localization methods:

- Relative localization where the robot position is computed by incrementing its preceding position and the measured variation with proprioceptive sensors. The two principal methods of relative localization are odometry and the inertial localization. These techniques use a non structured data and produce an accumulating error to estimate the robot position;
- Absolute localization requires the knowledge of the environment to determine exactly the robot position or to periodically readjust incremental estimate (navigation) produced with relative localization techniques. Exteroceptive sensors are used and various techniques can be distinguished to compute the robot position. The most known approaches are the magnetic compasses localization, the active reference marks localization, the passive reference marks localization and the model based localization techniques [3].

Vision techniques for autonomous mobile robotics are absolute and more specially model based localization techniques which put in correspondence perceived data with an initial model to determine the robot position. Models are built from exteroceptive sensors such as cameras or stereoscopic sensors. DeSouza [2] gather the existing approaches in three categories according to the a priori knowledge provided to the system and that the robot have to see:

- *Map-Based Localization (or Navigation).* The system uses a preset model designed by the human one.
- *Map-Building-Based Localization.* The system builds from its own sensors a model of its environment then uses it to the localization and navigation task.
- *Mapless Navigation.* The system does not use any explicit representation of the environment but is based on the tracking of recognized objects from the camera catches of sight.

This knowledge could be presented in a metric models form (geometrical primitives or occupation grids) or in a topological model form [3]. We propose in this paper a new approach for the robot localization problem which consists in using an image database model and consequently content based image retrieval techniques to provide a qualitative estimate of the robot position.

The remainder of this paper is organized as follows. In the next section we present our localization approach. Related works on content based image retrieval systems are given in section 3 and the data we used is presented in section 4. Color histograms are presented in section 5. the components and details of our system are described in sections 6 and 7 respectively. We present and discuss our results in sections 8 and 9 and we draw conclusions in section 10.

2. Localization approach

The central idea is to provide to the navigation system a set of images and features potentially visible and detectable by computer vision techniques during the navigation. The system's aim thus consists in searching primitives and features to identify the closest images from this set to determine the position and the orientation from which these images are seen.

This global localization is necessary after a long displacement of the robot to know its position whether it is lost and when the problem of the local localization is difficult to solve.

We work trough the ARPH project [4] (Robotics Assistance to Handicapped People) defined with the French Association against Myopathies (AFM). The aim of the project is to embark an arm manipulator (cf fig.1) 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.

Figure 1. *Handicapped person assistance robot*

For the global localization, using a color camera fixed in the base, we propose a content based image retrieval method. The principle is to constitute an image database of the house in which the robot evolves/moves. To find itself, the robot takes an image of its environment that we call request image. Then it seeks the image of the database which is closest to the query image and which is associated to the position and orientation's information.

With the difference to the majority retrieval systems, request images taken by the camera of the robot differ from images which constitute 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 nearest image from the one taken by the robot. This image will be used to determine the room where the robot is and its orientation in this room, two informations necessary for the global localization of a robot in an indoor environment. To achieve this goal, color is a necessary information to take into account. 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 descriptors which tolerate these changes, from which one can find the image in question, than complete descriptors which proves too restrictive and obliges to use not very stable quantities. What is required is the compactness of descriptors with the speed of computation since the image database is not very bulky.

3. Background

Content based image retrieval systems (CBIR) were essentially developed because of the quantities of digital images which are increasingly bulky. These images are, in general, compressed then filed in databases. Once these data stored, the problem is capacity to find them simply. The re-use of these databases passes by the joint development of methods of indexing and research. A coarse representation of such a data management can be described as follows:

$$\{\text{image}\} \rightarrow \text{descriptors} \rightarrow \text{indexing}$$

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

A content-based image retrieval system comprises generally four tasks. The two principal ones are obviously the indexation and research. The indexation consists in computing a signature summarizing contents of an image which will be then used for research. The attributes usually used within this framework are color, texture and shape. The research 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 localization problematic. The two other tasks are navigation and analyzes. Navigation is mainly related to the types of databases consultation. This functionality is often static with a search for one or more answers to a given request. A new type of research more interactive results in a more incremental approach and especially more adaptive to the needs of users. From the images results found at a first time search, the user can refine his research according to an object or a selected zone. In addition, the analysis is a very particular functionality which consists in providing quantitative results and not of visual nature (for example, the number of images with a blue color bottom). This functionality is thus summarized to extract statistics from characteristics of images.

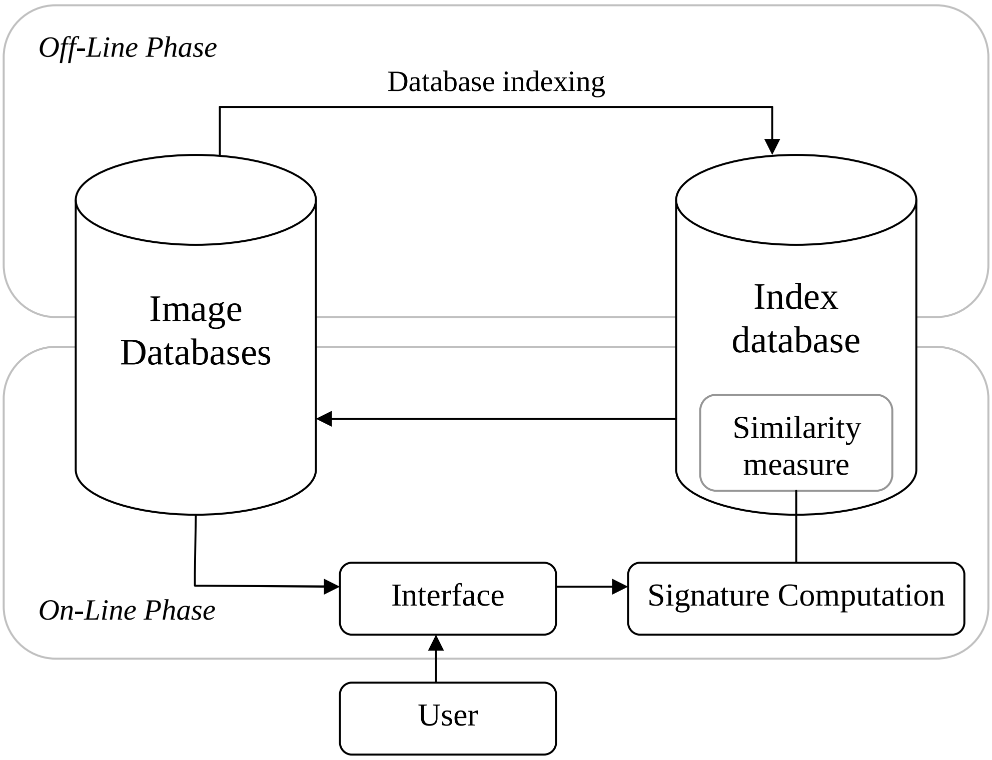


Figure 2. *Content based image retrieval architecture*

In addition, systems are generally based on a request by the example: further to a request image taken by a robot in our case, the search engine retrieves the closest images of the database within the meaning of a given similarity measure. The ideal tool of research is then which gives the quickly and the simply access to the required relevant images to a request image taken instantaneously by the mobile robot. It is thus a question of finding automatically, in the database, images visually similar to the request image. The similarity between two images is evaluated by using a specific criterion which can be based on color, shape, texture or a combination of these features. Many techniques were proposed with color based image retrieval [5, 6, 7], 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 [8]:

- constitution of a significant reference base allowing to store images and files of index associated with each image ;
- quantization of each image by keeping only the relevant colors in order to optimize the efficiency in time and in results ;
- images signatures development according to the desired requests (this signature consists of a combination of generic attributes and specific attributes related to the application) ;
- choice of a metric for the similarity measure;
- realization of an interface allowing requests by the example for the concerned applicability.

Many academic and/or industrial content-based image retrieval systems were developed: Mosaic [9], Qbic [10], Surfimage [11], Netra [12], VisualSEEK [13] ... They allow an automatic search for images per visual similarity. The standard architecture of all this marketed systems comprises an off-line phase to generate image indexes and an on-line phase for the research problematic (cf fig.2). Some systems are conceived for general public applications (for example the search for images on Internet). Image databases are then general and include heterogeneous 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 characterized by homogeneous contents (faces, medical images, fingerprints...). In the specific databases, the developed descriptors 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 descriptors are universal (color, texture, shape, etc.) [14]. Although our specific applicability (the global localization of a robot in an indoor environment), image databases are generic because of the variety of objects present in a house (cf fig.3).

4. Image databases

Two complete and well structured image databases are built for the global localization of the robot. The first database contains 240 images and the second 586 images. The size of images is 960x1280 pixels. These images have been taken in different rooms from two houses. 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. Figure 3 shows examples of images from the first database (a, b) and from the second one (c, d).

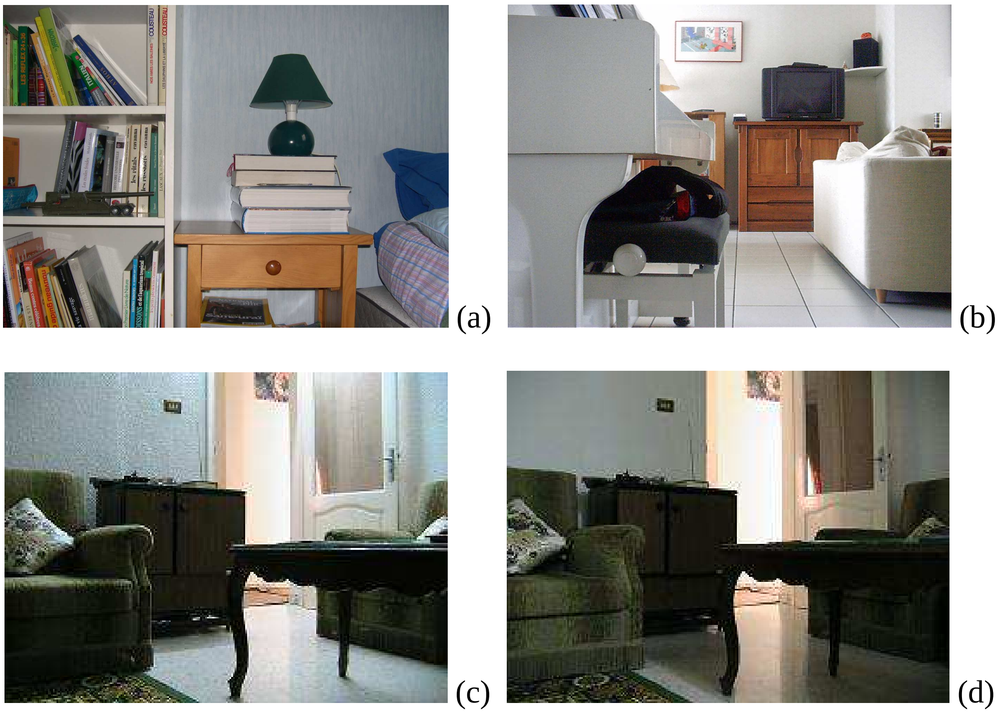


Figure 3. *Examples of indoor images*

In the second database we take also the luminosity into account (cf fig.3 (c, d)). For the same position, we have two or three images which have been taken at different day time. We took also 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.

5. Color Histograms

Colorimetric information is very significant in a domestic environment. Indeed, such a medium includes various elements without colorimetric coherence between them. A discrimination of these elements, and thus of images of the database, can be more powerful by taking into account their colors.

Color histograms remain the most used techniques as for the use of color information in retrieval systems. The robustness of this descriptor and its invariance to the position and orientation changes of objects make the strong points of it. Nevertheless, these performances are degraded quickly according to the size of the database. But in our application, the image database is not very bulky. Indeed, in an interior medium, we don't exceed a few hundreds of images to describe structurally the environment of the robot. The use of the histograms for color images indexing is based primarily on the techniques of selection of the adapted color space, the quantification of the selected space and the comparison methods by similarity measures. We have tested the RGB and the Luv color spaces. To the RGB color space which give best results, we developed several uniform quantizations with a goal of testing different size of pallets. Given a color image I , of size M by N pixels, the color distribution of a color bin $\boldsymbol{c}$ which ranges over all bins of the color space is given by:

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

In the equation above δ( ) is the unitary impulse function. We notice that the $h(\boldsymbol{c})$ values are normalized in order to sum to one. The value of each bin is thus the probability that the color $\boldsymbol{c}$ 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 non-corresponding bins (Mahalanobis distance and EMD Distance). Hereafter we present those similarity measures between a request image (I) and all the database images (H).

- **Minkowski distance :** $$d(I,H) = (\sum_k \left|h_k^I - h_k^H\right|^r)^{1/r} \qquad r \geq 1 \qquad (2)$$

Distance L_1 **:** r = 1
Distance L_2 : r = 2

- **Histogram intersection:** $$Inters\ (I, H) = \frac{\sum_k \min(\ h_k^I, h_k^H\ )}{\sum_k h_k^H} \qquad (3)$$

This function deducts the number of pixels of the model having a direct correspondent in the request image. Values close to 1 indicate a good resemblance [5].

- **The χ^2 test:** A color histogram can be considered as the realization of a random variable giving colors 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 [15]. The χ^2 test is given by:

$$\chi^2 = \sum_j \frac{(h_j^I - h_j^H)^2}{(h_j^I + h_j^H)^2} \qquad (4)$$

- **Mahalanobis distance** or generalized quadratic distance D_{QG} was used by Equitz and Niblack [26] to take into account the intercorrelation between color components. A weighting matrix W which include the resemblance between colors was proposed. The generalized quadratic distance resulting from the euclidean distance is defined by the following formula:

$$d_{QG}(I,H)=\sqrt{(H-I)W(H-I)^T} \qquad (5)$$

The components w_{ij} of the weighting matrix W can be interpreted like indices of similarity 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 color bins. Other proposals of weightings matrices attached to the representation of color spaces to define the colorimetric distances between colors were introduced recently [16].

- **EMD Distance** "Earth Mover Distance" proposed by Rubner [17] consists in extracting the minimal quantity of energy necessary to transform a signature into another. Having the distances d_{ij} between colors components of the two histograms H and I of m and n dimensions respectively, it is a question of finding a whole of flow F = [f_{ij}] which minimizes the cost of the following quantity:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij}\ f_{ij} \qquad (6)$$

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 color component should be defined. From the whole of optimal displacements F, EMD distance is then defined like the following resulting work standardized:

$$d_{EMD}(H,I)=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \quad (7)$$

The formalism suggested by Rubner meets all conditions to determine the optimal distance between two histograms but the complexity introduced by the algorithm of optimization makes it expensive in computing time [18].

6. A new color feature definition

6.1. Baker’s transformation

The baker's transform (BT for short) is based on the definition of mixing dynamical systems [19, 20]. The main interest of these transformations is that they mix in a very homogeneous way all the elements of the involved space.

Arnold & Avez [19] 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 & 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, when restricted to finite sets, it remains some mixing like properties: the pixels are statistically well mixed by a suitable number of the BT’s iteration.

One iteration of the BT is based on two steps: first, an “affine” transformation is used which gives an image twice larger and half high (cf. fig 5) from an original image (cf. fig 4). Then, the resulting image is cut vertically in the middle and the right half is put on the left half (cf. fig 6). After a suitable number of iterations, we obtain a well mixed image (cf. fig 7). From this mixed image we extract a definite size window which gives after some iterations a reduced scale version of the original image (cf. fig 8). The BT requires that the image size is 2^Nx2^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^px2^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^px2^p window.

Figure 4. *256x256 original image*

Figure 5. *First step of BT initial iteration*

Figure 6. *Second step of BT initial iteration*

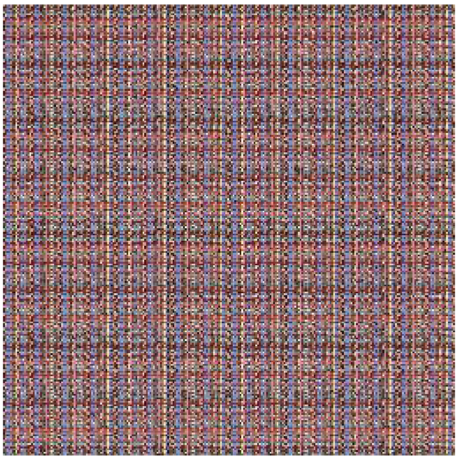


Figure 7. *Well mixed image*

Figure 8. *16x16 pallet deduced from the mixed window*

6.2. The color feature

As shown in figure 8, a little image of size 16x16 gives a good color, shape and texture representation of the original image and we can consider it as a representative color pallet. In [22], we presented a first use of this method to quantify color images. The idea is to use one of these windows as a color pallet to reduce all the color levels of the original image. With a $2^N x 2^N$ image, it is possible to propose pallets containing 2^{2P} colors (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 color and all colors present in the pallet. Then the new color assigned to the pixel is that which minimizes the distance. The problem is "how to choose the representative window to build the good pallet?". We analyze 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 distance computing between the original image and the reduced one. This distance, baptized "delta", is computed on each of the three color channels (Red, Green and Blue) for all image pixels, by using equation 8, where I_1 and I_2 respectively represent the color levels of a pixel in the initial image and the reduced one.

$$delta = \frac{\sum_{i=1}^{2^N} \sum_{j=1}^{2^N} \left| I_1(i,j) - I_2(i,j) \right|}{2^N \times 2^N} \tag{8}$$

From a practical point of view, BT is a space transformation. For a given dimension of image, the position of the arrival pixels in the mixed image is always the same one. Consequently, a Look Up Table (LUT), indicating for each pixel of an image its co-ordinates in the mixed image, will allow obtaining the pallet more quickly. In addition, the BT or the use of a LUT amounts extracting in a homogeneous way pixels from the image. Thus, it is possible, for rectangular images, to obtain a same feature by applying a sub sampling technique.

7. Retrieval approaches

7.1. Color reduction retrieval approach

If it is possible to extract a sample of pixels, which the colors are representative of the original image and which are stable for images having the same sight, then we can call this sample a color invariant. This color feature is used as an indirect descriptor [23]. The strategy to search the closest image from the database, to the request image, is shown on figure 9. First we build a pallet database by computation of the color invariant for each image from the original database. Then, the request image is projected in the color space defined by each pallet from this pallet database. We compute the color difference between the request image and the projected images (cf. table 1) and we select the pallet (i.e. the image) which leads to the minimum of this distance.

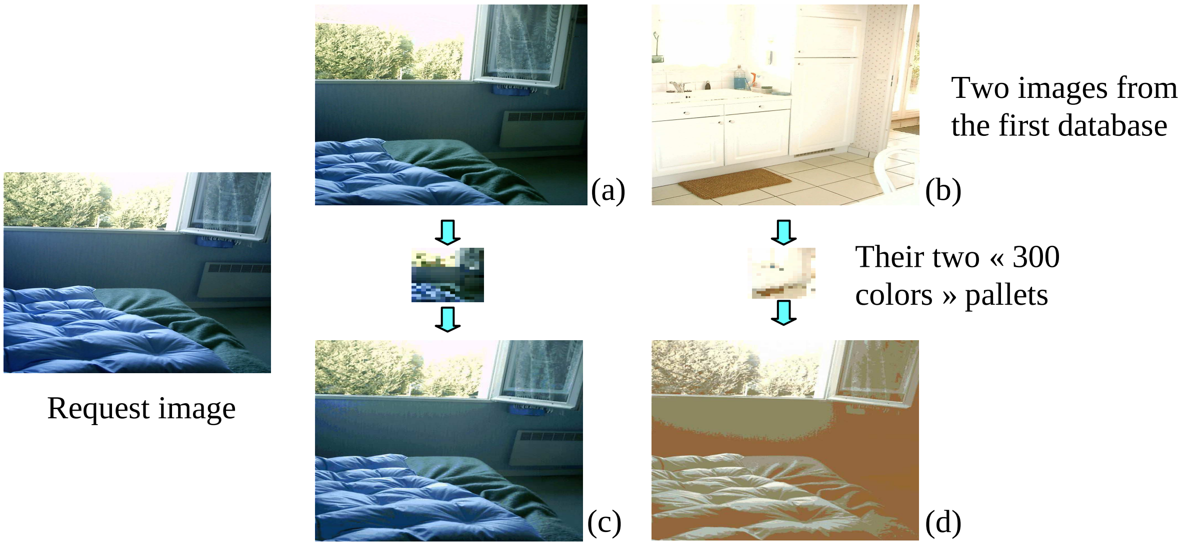


Figure 9. *Request image reduced by pallets of the images (a) and (b) give the result images (c) and (d) respectively*

Table 1. Differences between request image and reduced ones

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

7.1.1. Results of the color reduction retrieval approach

From the two image databases, we have build 2x5 pallet databases, to test five number of colors: 48, 108, 192, 300 and 588, which respectively correspond to pallets of: 6x8, 9x12, 12x16, 15x20 and 21x28. In order to speed up the retrieval process, we sub sampled the request image (60x80 pixels). Tables 2 & 3 show synthesis of the obtained results. The answers are organized in three classes:

“Right”: The image proposed by the retrieval system has been taken in the same room and with the same orientation than the request image.

“Medium”: The image proposed by the retrieval system has been taken in the same room than the request image.

“False”: The image proposed by the retrieval system has been taken in other room than the request image.

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

	Color number	48	108	192	300	588	%
First answer	*Right*	5	9	8	9	9	**40,0**
	Medium	6	3	4	4	2	**19,0**
	False	9	8	8	7	9	**41,0**
Three answers	*Right*	10	11	13	13	13	**20,0**
	Medium	24	21	17	18	21	**33,7**
	False	26	28	30	29	26	**46,3**

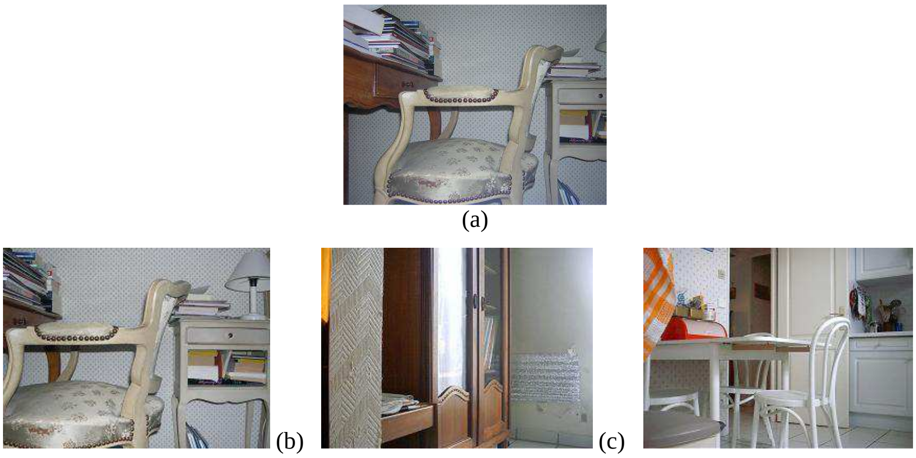
Table 3. Results for database n°2 – 35 request images

	Color 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**

We analyzed two cases: the quality of the first answer and the quality of the three first answers. We can see we obtain 40% or more of good answers when we take one answer only into account. If we want a coarse answer to the question: In which room is the robot?, we sum the classes "good" and "medium". Then the rate of correct answer is about 60% for the database n°1 and over 70% for the other. When we take the first three answers into account, we obtain bad results. Then, we will work with the first answer only.

Moreover, if a poor number of colors in the pallet leads to a bad result, a great number of colors doesn't give a better result. In database n°2, we obtain results over 75% with 192 and 300 colors in the pallet. We retain this last size (300 colors) to work with for the next experiments.

Figures 10.a and 11.a show request images from the first and the second databases respectively. Figures 10 (b, c and d) present the three answers obtained (10b give the right response, 10c and 10d are false). Figures 11.b and 11.c present two examples of the first answer obtained with two different pallets. We can see that the result is right with 192 colors (11b), but it is false with 48 colors (11c).



(a)

(b) (c) (d)

Fig. 10. *Three answers with a pallet of 300 colors from the request image (a)*

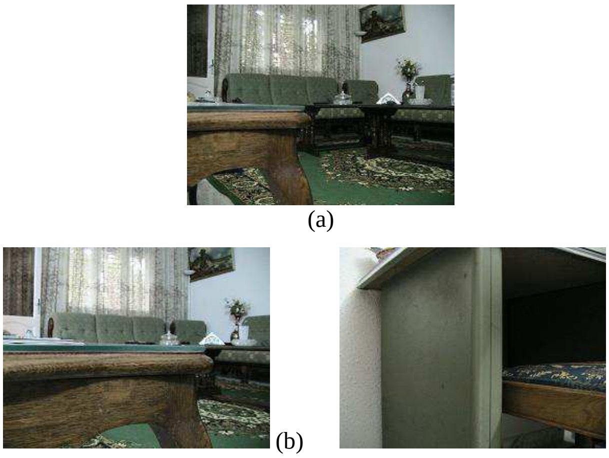

(a)
(b) (c)

Fig. 11. *First answer with a pallet of 192 colors (b) and 48 colors (c) from the request image (a)*

In spite of its interest which validates the concept of color 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 time that the database is bulky. We can however consider the pallet as a feature and compare pallets between them in the search phase instead of comparing request image with reduced ones.

7.2. The interpallet distance

After a first use of this color pallet as an indirect descriptor, we associate to this feature an Euclidean distance that we call interpallet distance $L_2(P_{req} - P_{base})$ [24]. The strategy to search the closest image, from the database, to the request image, is described as follows (cf. fig 12).

- First we build a pallet database by the computation of the color invariant for each image from the original database.
- Then, we extract the pallet of the request image to compute the color difference between this one and all pallets already built in the database. Euclidean distance is computed between correspondent color having the same position in the pallets.
- Finally, we select the pallet (i.e. the image) which leads to the minimum of this distance.

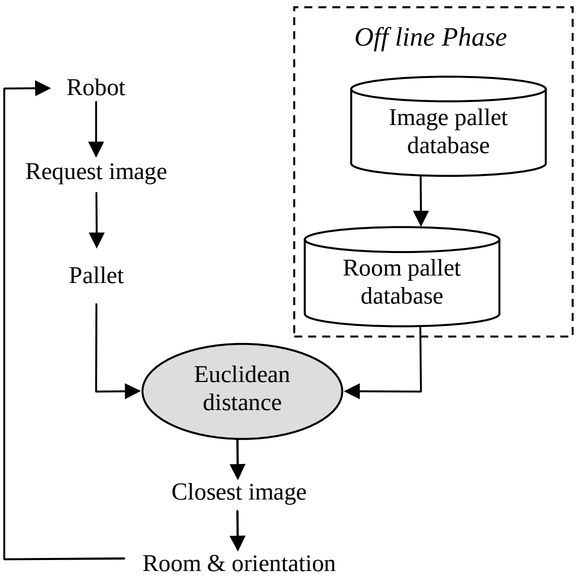


Figure 12. *Interpallet distance.*

The space organization of colors of this two dimensional pallet is an additional information who can present invariance property to some changes in image sample. Thus, we emphasis this color feature aspect and try to model it by preserving the interpallet distance who give an interesting results. Indeed, as the below figure shows it, the pallet preserves the spatial distribution and the principal vicinity relations between colors 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).

7.3. Space distribution of colors

To describe the space aspect of the pallet, we built its colors statistical moments in order to coarsely describe colors distribution form of the image. Stricker [16] establishes a balanced sum of the average, the variance and skewness (the third order moment) computed for each color channel, to provide a single number used in the indexing process. If p_{ij} is the value of the pixel j for channel i, these moments are defined by:

$$\mu_i=\tfrac{1}{N}\sum_{j=1}^{N}p_{ij} \quad , \quad \sigma_i=\tfrac{1}{N}\sqrt{\sum_{j=1}^{N}(p_{ij}-\mu_i)^2} \quad \text{and} \quad s_i=\tfrac{1}{N}\left(\sum_{j=1}^{N}(p_{ij}-\mu_i)^3\right)^{\frac{1}{3}} \tag{9}$$

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

$$d_{mom}(I,H)=\sum_{i=1}^{3}w_{i1}\left|\mu_i^I-\mu_i^H\right|+w_{i2}\left|\sigma_i^I-\sigma_i^H\right|+w_{i3}\left|s_i^I-s_i^H\right| \tag{10}$$

We have applied these moments on our two dimensional pallet. p_{ij} are in this case pixels from the pallet. We notice that a space description of our two dimensional pallet by color moments as showed in [13], gives better results than a similar description of the entire original image. We deduce that this description of a pallet which represent on a reduced scale its 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 1260x960 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 our interpallet distance. We compute the variation of these first four moments. The variance which has the greatest dynamics 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.

$$w_1=\lambda\,\frac{\Delta\sigma}{\sigma_{im}+\sigma_{req}} \tag{11}$$

with

$$\Delta\sigma=\begin{cases}\alpha & \text{if } \left|\sigma_{req}-\sigma_{im}\right|<\alpha\\ \left|\sigma_{req}-\sigma_{im}\right| & \text{else}\end{cases} \tag{12}$$

and

$$\lambda=\begin{cases}1 & \text{if } \left|\sigma_{req}-\sigma_{im}\right|<\beta \; and \; \left|\mu_{req}-\mu_{im}\right|<\beta\\ \infty & \text{else}\end{cases} \tag{13}$$

Thus

$$D_1 = w_1 \,.\, L_2(P_{req} - P_{im}) \tag{14}$$

7.4. Vicinity template of colors

To describe the textural aspect of colors distribution, we developed the co-occurence matrix and some relating features defined by Haralick [25] and extended to color information by Tremeau [26] which are:

- Color Inertia: $$I = \sum_{i=0}^{M} \sum_{j=0}^{M} D_{ij}^2 . P(i,j) \tag{15}$$

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 color channels of the RGB color space.

- Color correlation: $$C = \sum_{i=0}^{M} \sum_{j=0}^{M} \frac{D_i . D_j}{\sigma_i . \sigma_j} P(i,j) \tag{16}$$

with $D_i = ((R_i - R_{\mu_i})^2 + (G_i - G_{\mu_i})^2 + (B_i - B_{\mu_i})^2)^{1/2}$

and $D_j = ((R_j - R_{\mu_j})^2 + (G_j - G_{\mu_j})^2 + (B_j - B_{\mu_j})^2)^{1/2}$

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

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

with $$R_{\mu_i} = \frac{1}{\sum_{j=0}^{M} P(i,j)} . \sum_{j=0}^{M} P(i,j) . R_j \tag{17}$$

and $$\sigma_i = \sqrt{\frac{1}{\sum_{j=0}^{M} P(i,j)} . \sum_{j=0}^{M} P(i,j) . D_j^2} \tag{18}$$

- homogeneity : $$H = \sum_{i=0}^{M} \sum_{j=0}^{M} P(i,j)^2 \tag{19}$$

- Entropy : $$E = \sum_{i=0}^{M} \sum_{j=0}^{M} P(i,j) . \log P(i,j) \tag{20}$$

Moreover, we extract the maximum value of the co-occurence matrix and its two color components that we note (c_1, c_2).

Owing to the fact that a fine quantization of a color space gives a substantial signature, the construction of co-occurence matrices of pallets (low-size images) bring smooth and not enough discriminating distributions. To mitigate this problem with modelling the main colors vicinity, we developed the co-occurence matrices with a coarse uniform quantization of the RGB color space in 64 bins. We considered, in addition, an isotropic vicinity (8 connexities) of each pixel.

For the search phase, we developed the Euclidean distance $L_2(M_{req} - M_{im})$ between co-occurence matrices of pallets. We preceded this distance by a weighting factor w_2 especially built from the entropy variation which has the greatest dynamics and so the discriminating capacity among the other co-occurrence matrix features what gives the D_2 distance hereafter:

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

with

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

thus

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

We estimated the value λ according to the three dimensional connexity of the color components of the co-occurence matrix maximum value between the request image and each database image. By assimilating uniform color bins to cubes (cf. fig 13), the three dimentional connexity is evaluated as follows:

Connex1: cubes adjacent by a surface e.g. cubes (1,2), (1,3)

Connex2: cubes adjacent by an edge e.g. cubes (2,3), (3,5)

Connex3: cubes adjacent by a point e.g. cubes (2,4).

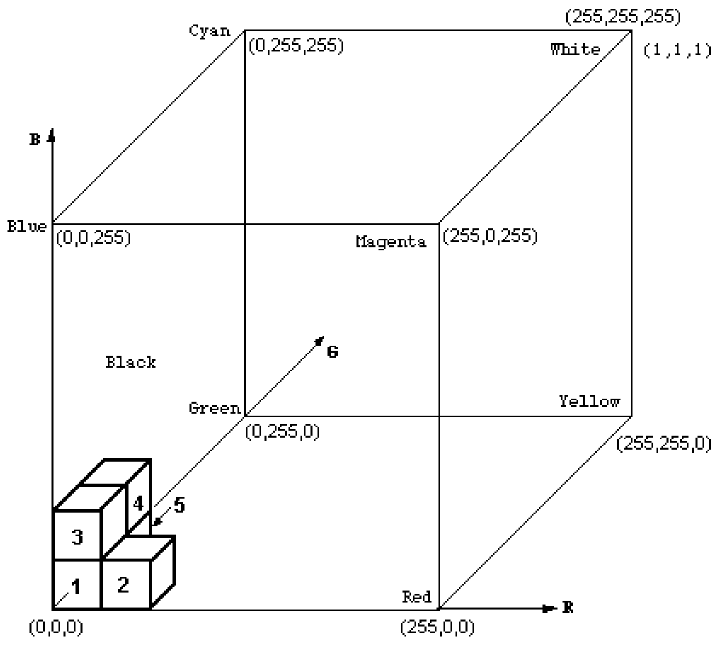


Figure 13. *RGB color cube*

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

If $(c_1,c_2)_{req} = (c_1,c_2)_{im}$ Then $\lambda = 1$
If $(c_1,c_2)_{req}$ Connex1 $(c_1,c_2)_{im}$ Then $\lambda = 2$
If $(c_1,c_2)_{req}$ Connex2 $(c_1,c_2)_{im}$ Then $\lambda = 3$
If $(c_1,c_2)_{req}$ Connex3 $(c_1,c_2)_{im}$ Then $\lambda = 4$
Else $\lambda = 8$

7.5. The final distance D

The D_2 distance built by co-occurence matrices of the pallets gives lower results than the D_1 distance (only 55% of right room). But by finely analyzing 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 normalized distances between pallets and between co-occurence matrices into account, each one balanced by a resulting term from color moments and co-occurence matrices attributes respectively. D is given by:

$$D = w_1 \, L_2(P_{req} - P_{im}) + w_2 \, L_2(M_{req} - M_{im}) \quad (24)$$

7.6. Hierarchical approach

We conceive, in a preliminary stage, before applying the proposed distance *D*, a hierarchical search using classification of images according to rooms. We characterize each room by a discriminating color pallet. This room pallet is built by sampling color pallets of images taken in this room and by adding colors whose minimal distance to those retained is higher than a threshold fixed at 10.

In the search phase, we compute the difference between the request pallet and room pallets built previously through the Euclidean distance and we classify these distances by ascending order. We eliminate firstly rooms presenting non similar colors to those of the request image (cf fig. 14). Finally, we apply the *D* distance to images from the database taken in the rooms where the robot is most probably lost.

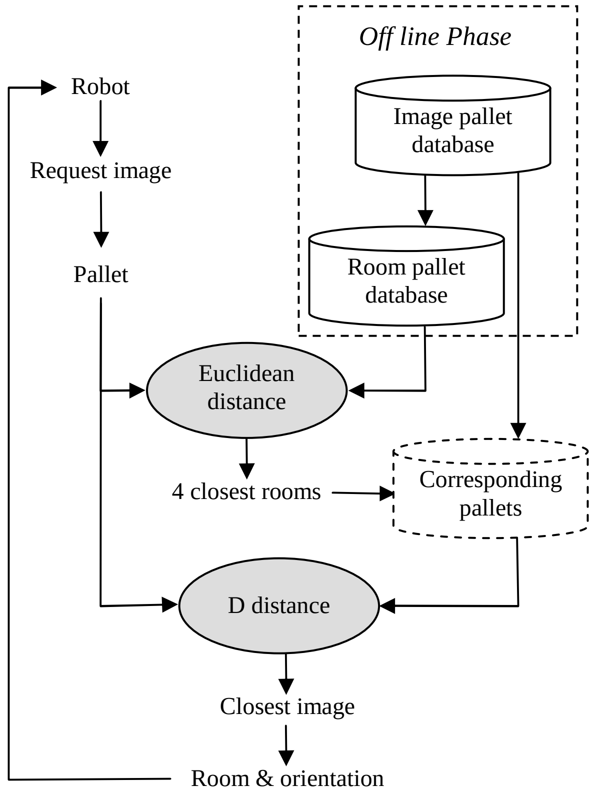


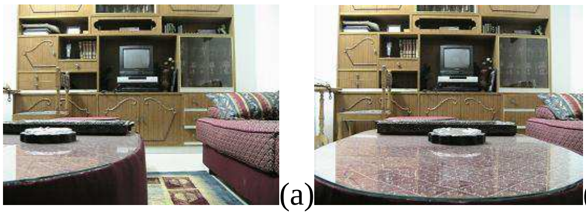
Figure 14. *Hierarchical search.*

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 research the first four rooms given by the hierarchical process. It should be noted that our aim is to eliminate rooms on which certain images distort our results. This, being made, the performances of our system are improved.

8. 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 organization and color vicinity as the colorimetric aspect of the pallet. For the request image (cf fig. 15.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 the figure 15-b indicating the right room and the right orientation.



(a) (b)

Figure 15. *a) Request Image from the second database;*
b) Response Image within D distance.

Table 4 shows results of our retrieval strategies based on a two dimensional color pallet extract from the Baker's transformation which show the interest of our approach.

Table 4:Results of our methods

	Database 1				Database 2				Databases1 & 2
	Time	Results	%		Time	Results	%		%
	Color reduction retrieval approach								
Right	0.7 sec	9	45%	65%	1.1 sec	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 sec	10	50%	60%	1.1 sec	15	42.8%	60%	60%
Medium		2	10%			6	17.2%		
False		8	40%			14	40%		40%
	Spatial and color distance D								
Right	2 sec	12	60%	65%	3 sec	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 sec	12	60%	70%	5 sec	20	57%	88,5%	82%
Medium		2	10%			11	31,5%		
False		6	30%			4	11,5%		18%

The first global retrieval approach based on the color reduction principle, while giving acceptable results (about 70%) proved limited by a prohibitory computing time. We developed, thereafter, a method by our color pallet description with an interpallet distance.

Results were a little degraded while reducing the computing time appreciably. Seeking to optimize the quality of description as well as the computing time, we took into account the space organization of the pallet to define a specific new similarity measure. Thus we could improve the results with a search time of about three seconds.
In order to improve even more the performances of room identification, we developed a hierarchical research method allowing to eliminate in the beginning a certain number of room from research to combine speed and effectiveness. The results are clearly improved to exceed the 80%. The hierarchical procedure being consuming in calculation 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 localization.

In order to validate this work, we compare these results with a classical image retrieval technique which uses color histogram. We developed color histograms on RGB and Luv spaces. The RGB color space which gives best results is performed to three uniform quantization into 64, 512 and 4096 color 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.

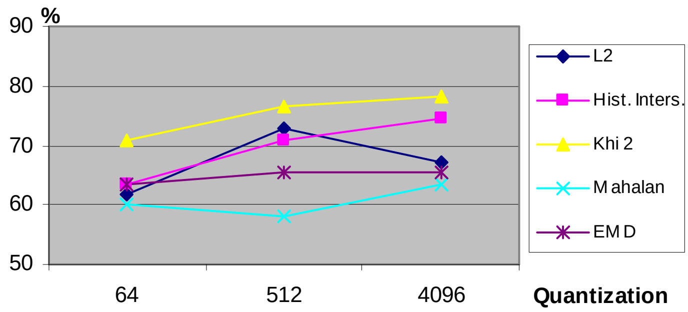


Figure 16. *Histogram results*

As showed in figure 16, the quantization to 64 colors proves very coarse for color histograms. Quantization to 512 bins improve considerably these results but the 4096 bins discretization gives the best results except the Euclidean distance which give best results with 512 bin quantization. We explicit results of the 4096 color histograms in the table 5. We note the worst results with the cross-bin similarity measures which tend to overestimate the mutual similarity of color distributions. Moreover, the computational complexity of the EMD and Mahalanobis distance are the highest among the evaluated measures. Indeed, computing the EMD between 4096 color histograms in our database take 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 localization task.

Table 5: Histogram results

	Base 1 & 2			
	Time	Results		
	Histogram Intersection			
Right	4 sec	31	56,4%	74,5%
Medium		10	18,1%	
False		14	25,5%	
	Euclidean Distance L_2			
Right	4 sec	27	49,1%	67,2%
Medium		10	18,1%	
False		18	32,8%	
	χ^2 test			
Right	4 sec	32	58,2%	78,2%
Medium		11	20%	
False		12	21,8%	
	Mahalanobis Distance			
Right	60 sec	20	36,4%	63,6%
Medium		15	27,2%	
False		20	36,4%	
	Earth Mover Distance : EMD			
Right	35 mn	23	41,8%	65,4%
Medium		13	23,6%	
False		19	34,6%	

Our method gives better results than those of the color 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. This results provided only with a color based description of indoor images is an encouraging result. A final system obviously must integrate other type of signature like shape and texture with a more structured model of the environment.

9. Further research

We can identify the following avenues for improving performance:

1. A first prospect for this work is to develop a local search approach to images. A combination of the global solution and the local approach has to improve even more the performances of our system. The characteristics developed in this paper were computed broadly in the image. However, a system only based on global characteristics cannot give the desired results. Indeed, an image made up of several objects having very different characteristics (colors and textures), the feature vector extracts from the whole image loses local information (the objects) and produces only one coarse average of the contents of this image. A possible solution consists to index some known and non removable objects in the house. A preliminary research phase could determine whether one of these objects is in the sight field of the robot reducing the complexity of the global search. 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 color distribution for our similarity measure e.g. by using Tamura signature features [27] like directivity, contrast and coarseness with a more fine

color pallet, or a frequency based model [28] can introduce texture useful information to improve discrimination between images.

3. Another prospect for research would consist on the exploration and the search for other descriptors and invariants such as differential invariants for color images and invariants for predictable change of illumination [29]. 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.

10. Conclusion

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

References:

[1] J.-J. Gonzalez-Barbosa, S. Lacroix, "*Localisation d'un robot mobile dans un environnement naturel par indexation d'images panoramiques*", LAAS/RIA Seminars, 2001.

[2] G. N. DeSouza, A. C. Kak, "*Vision for Mobile Robot Navigation: A Survey*", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 2, 2002, pp. 237-267.

[3] O. A.-Aider, "*Localisation référencée modèle d'un robot mobile d'intérieur*", these, 2002.

[4] P. Hoppenot, E. Colle, O. Ait Aider, Y. Rybarczyk, "*ARPH - Assistant Robot for Handicapped People - A pluridisciplinary project*", *IEEE Roman'2001*, France, 2001, pp. 624-629.

[5] M. Swain, D. Ballard, "*Color indexing*", International Journal of Computer Vision, 7 :1, 1991,pp 11-32.

[6] Y. Rui, T. S. Huang, "*Image Retrieval: Current Techniques, Promising Directions, and Open Issues*", Journal of Visual Communication and Image Representation 10, 1999,pp 39–62.

[7] J.R. Smith and S.-F. Chang, "*Local color and texture extraction and spatial query*". In Proc. IEEE Int. Conf.On Image Processing, vol. 3, 1996., pp. 1011--1014,

[8] C. F.-Maloigne, "*Quantification et segmentation pour l'indexation de bases d'images couleur*", TAIMA, 2001, pp 61-66.

[9] S.-T. Goh, K.-L. Tan, "*MOSAIC: A fast multi-feature image retrieval system*"**,** Data & Knowledge Engineering, Volume 33, Issue 3, June 2000, pp. 219-239.

[10] M. Flickner, "*Query by Image and Video Content: The QBIC System*", IEEE Computer, vol. 28, no. 9, 1995, pp. 23-32.

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

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

[13] J.R. Smith and S.-F. Chang, "*Visualseek: A fully automated content-based image query system*", In Proceedings of ACM Multimedia, 1996, pp. 87-98.

[14] N. Boujemaa, S. Boughorbel, V. Constantin, "*Description de la répartition spatiale de la couleur pour l'Indexation d'Images*" RFIA, vol. 2, 2002, pp 405-414.

[15] P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid, G. Mistler, "Utilisation *de la couleur pour l'appariement et l'indexation d'images*", INRIA, Report N° 3269, Topic 3, 1997.

[16] M. A. Striker, M. Orengo, "*Similarity of color images*", SPIE Conference on Storage and Retrieval for Image and Video Databases V, vol. SPIE-2420, 1995, pp. 381-392.

[17] Y. Rubner, C. Tomasi, L. J. Guibas, "*A metric for distributions with applications to image databases*", In IEEE International Conference on Computer Vision, 1998, pp. 59-66.

[18] A. Trémeau, C. F.-Maloigne, P. Bonton, "*Image numérique couleur*", France, Dunod, 2004, 459 pages.

[19] V. I. Arnold and A. Avez, "*Problème ergodiques de la mécanique classique*", Monographies Internationales de Mathématiques Modernes n°9, Gauthier-Villars, Paris, 1967

[20] P. Billingsley, "*Ergodic Theory and Information*", John Wiley & Sons, Inc., New-York, 1965

[21] A. Smolarz and P. Cornu, "*A Color Image Classification by Means of Image Transformations*", CGIV'2002, Poitiers, France, 2002, pp. 310–315.

[22] C. Montagne, S. Lelandais, A. Smolarz and P. Cornu, "*Adaptive Color Quantization Using the Baker's Transform*", CGIV'2004, Aachen, Germany, April 2004, pp. 353-358.

[23] C. Montagne, S. Lelandais, A. Chaari and M. Ben Ahmed, "*Invariant couleur basé sur la transformation du boulanger – Application à la localisation globale d'un robot d'intérieur*", SETIT, Tunisia, 2005.

[24] S. Lelandais, A. Chaari, A. Smolarz, C. Montagne, B. Jacquin, "*A new color invariant for image retrieval using the baker's transformation*", BICI, China, 2005.

[25] R. M. Haralick, K. Shanmugan, I. Dinsytein, "*Textural features for image classification*", IEEE Trans. Syst. Man & Cybern., vol. SMC-3, n°6, 1973, pp 610-621.

[26] A. Trémeau, B. Gérin, "*Utilisation de la matrice de cooccurrence couleur pour l'analyse de texture*", Automatic Conference. Data-processing genius. Image, 1996, pp. 303-306.

[27] H. Tamura, S. Mori, and T. Yamawaki, "*Texture features corresponding to visual perception*", *IEEE Trans. On Sys., Man. and Cyb.* SMC-8(6), 1978, pp. 460 - 473.

[28] K. S. Thyagarajan, T. Nguyen, and C. Persons, "*A maximum likelihood approach to texture classification using wavelet transform*", in Proc. IEEE Int. Conf. on Image Proc., Vol. 94, 1994, pp. 640 - 644.

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