



HAL
open science

An automatic natural feature selection system for indoor tracking - application to Alzheimer patient support

Mohamed Badeche, Frédéric Bousefsaf, Abdelhak Moussaoui, Mohamed Benmohammed, Alain Pruski

► **To cite this version:**

Mohamed Badeche, Frédéric Bousefsaf, Abdelhak Moussaoui, Mohamed Benmohammed, Alain Pruski. An automatic natural feature selection system for indoor tracking - application to Alzheimer patient support. *International Journal of Computational Vision and Robotics*, 2018, 8 (2), pp.201-210. 10.1504/IJCVR.2018.091982 . hal-02931121

HAL Id: hal-02931121

<https://hal.science/hal-02931121>

Submitted on 18 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An automatic natural feature selection system for indoor tracking – application to Alzheimer patient support

Mohamed Badeche*

Lire Laboratory,
Constantine 2 – Abdelhamid Mehri University,
BP 325, Route Ain El Bey,
25017, Constantine, Algeria
Email: badeche_mohamed@yahoo.fr
*Corresponding author

Frédéric Bousefsaf

Vision and Content Engineering Laboratory – CEA LIST,
CEA SACLAY – NANO INNOV,
Bât. 861 – Point courrier 173,
91191 Gif Sur Yvette Cedex, France
Email: frederic.bousefsaf@cea.fr

Abdelhak Moussaoui

Lcoms,
University of Lorraine,
7, rue Marconi, 57070,
Metz, France
Email: abdelhak.moussaoui@gmail.com

Mohamed Benmohammed

Lire Laboratory,
Constantine 2 – Abdelhamid Mehri University,
BP 325, Route Ain El Bey,
25017, Constantine, Algeria
Email: ben_moh123@yahoo.com

Alain Pruski

Lcoms,
University of Lorraine,
7, rue Marconi, 57070,
Metz, France
Email: alain.pruski@univ-lorraine.fr

Abstract: In this paper, we propose an automatic selection and natural feature tracking method that uses a monocular camera for path capturing and guides the user showing him the path to be followed. The application targets Alzheimer patients for helping them in their indoor moves. By offering an automatic selection of features, the user intervention and prior knowledge of the working environment would not be required to assure the good working of the system. The general principle of the proposed method is to record the path to be followed, and then recognise it in real time using purely visual methods, using only a single camera as an acquisition sensor. The devised system could be implemented on augmented-reality glasses with one single built-in camera. The experimental results have shown that the proposed method is very promising and the application could follow accurately the required path in real time, with a satisfying robustness in a fully-contrasted and static environment.

Keywords: natural features; matching; local descriptor; optical flow; Alzheimer disease; augmented reality glasses.

Reference to this paper should be made as follows: Badeche, M., Bousefsaf, F., Moussaoui, A., Benmohammed, M. and Pruski, A. (2018) 'An automatic natural feature selection system for indoor tracking – application to Alzheimer patient support', *Int. J. Computational Vision and Robotics*, Vol. 8, No. 2, pp.201–220.

Biographical notes: Mohamed Badeche is a PhD student at Constantine 2 – Abdelhamid Mehri University, Algeria. He received his Engineering and Magister degrees from Mentouri University and University of Batna in 2003 and 2006, respectively. His research interests include computer vision and augmented reality.

Frédéric Bousefsaf is a Postdoctoral Researcher Fellow at the French Alternative Energies and Atomic Energy (CEA), France. He received his PhD degree from University of Lorraine, France, in 2014. His research interests include computer vision, signal processing, biomedical engineering and affective computing.

Abdelhak Moussaoui is an Associate Professor at University of Tlemcen, Algeria. He received his PhD from University of Metz, France, in 2010. His research interests include virtual reality, exposure therapy and social phobia.

Mohamed Benmohammed is a Full Professor of Computer Science at Constantine 2 – Abdelhamid Mehri University, Algeria. He received his PhD from University of Sidi Bel-Abbes, Algeria, in 1997. His research interests include microprocessor, computer architecture, embedded system and computer networks.

Alain Pruski is a Full Professor at Fundamental and Applied Science Department of University of Lorraine, France. He received his PhD from University of Lorraine, France, in 1984. His research interests include virtual reality and cognitive stimulation.

1 Introduction

According to the World Health Organization, the Alzheimer's disease affects between 21 and 25 million people around the world (WHO, 2012), and has heavy social and human consequences for both the patient and his family (Thomas et al., 2006). While medical advances are trying to heal diseases, technological advances allow the patient to interact with the world. The work presented in this paper offers a technological support to help Alzheimer patients (APs) who are not able to properly remember their itineraries.

Amnesia is one of the most common APs' problems. Actually, it is one of the first symptoms used to properly diagnose the disease (American Psychiatric Association, 2013; Dubois et al., 2007). An AP with memory loss issues would need someone to escort him in his indoor moves, since memory loss of familiar places appears at the earliest stage of the disease (WHO, 2012). Thus, the patient is likely to face serious problems to reach a specific place or to perform a given task.

The disease evolution often evolves through seven stages. Before the fifth stage, the patient is relatively autonomous in accomplishing his daily tasks (Reisberg and Franssen, 1999), the problem here is that he is not unable to accomplish a given task, but he would get real problems in reaching the place of that task, for example an AP would need, before the fifth stage, someone to escort him to the rest room, but once he reaches it, he would need no further assistance.

In this context, this work proposes an automatic path tracking method, mainly destined for the APs, in order to compensate their difficulty to remember their itineraries. Tests have been conducted on computers, but the final purpose of our work concerns the implementation of the method on augmented-reality glasses; in order to guide the patient through a given path, such as going from the bedroom to the rest room, without needing any intervention from a caregiver. The devised system recognises the path and locates the patient without any prior modification of the environment, it does not ask him to modify the environment, by adding any markers, and without using another sensor except the camera itself.

In the literature, prior knowledge of the environment has been used to give proper indications to the user. For example, Hile and Borriello (2008) employed alignment of the frames on the basis of the building floor plan to give the user positioning and orientation. Quintana and Favela (2013) used a system of augmented reality annotations on objects, previously defined by the caregiver, to assist the patient in his moves.

Our method, thus, differs from previous works, like those proposed in Bai et al. (2014) and Nozawa et al. (2012), in two aspects; on the one hand, the selection of features in our proposal is automatic and does not require any user intervention, and on the other hand, it is mainly oriented to APs, it helps them by showing them directions on the augmented reality glasses.

Orienting the patient is done using direction arrows, which mainly indicate three directions: straight, as seen from Figure 1, right or left. These guidelines, which are displayed in real time, would be sufficient to guide the user in his moves in a flat environment and even in an environment with slopes and stairs. The choice to take a slope or a descent by being close enough may mean having the choice to go right, left or straight on.

Obviously, other complex displaying methods can be used but overloading the glasses with too much information might disturb the user. However the display of simple direction arrows can be both beneficial and easily assimilated by the user.

Even in the presence of amnesic troubles, the AP at an early stage can autonomously avoid obstacles. Thus; the proposed method does not take into account avoiding dynamic obstacles, which do not constitute any difficulty to the patient; the AP forgets his path but remains aware of how to avoid obstacles.

Figure 1 Direction guidelines (see online version for colours)



The rest of the paper is organised as follows: Section 2 presents some basic concepts related to image matching techniques and perspective projection. The proposed approach is given and detailed in Section 3. Some experimental results are discussed in Section 4, and we conclude the paper in Section 5.

2 Basic concepts

In this section, some concepts related to image matching techniques and perspective projection are presented.

2.1 Matching techniques

Since we aim at proposing a visual path tracking approach, it is a necessity to apply an image recognition step on the input video sequence. To do so, we opted for using a technique based on local descriptors, which proved to be the most suitable matching techniques (Tuytelaars and Mikolajczyk, 2008), namely the SURF technique (Bay et al., 2008).

As far as local descriptors are concerned, many comparative studies have been proposed in the literature. We cite two of them here. Mikolajczyk and Schmid (2005) have compared shape context (Belongie et al., 2002), steerable filters (Freeman and Adelson, 1991), principal components analysis – scale invariant feature transform (PCA-SIFT) (Ke and Sukthankar, 2004), differential invariants (Koenderink and van Doorn, 1987), spin images (Lazebnik et al., 2003), scale-invariant feature transform

(SIFT) (Lowe, 1999), complex filters (Schaffalitzky and Zisserman, 2002) and, finally, the gradient location and orientation histogram (GLOH), proposed by the same authors as an extension of SIFT in the same paper. Fu (2014) has proposed, in another recent work, a comparison between: SIFT, speeded up robust features (SURF), maximally stable extremal regions (MSER) (Matas et al., 2004), binary robust independent elementary features (BRIEF) (Calonder et al., 2012), binary robust invariant scalable keypoints (BRISK) (Leutenegger et al., 2011), oriented FAST and rotated BRIEF (ORB) (Rublee et al., 2011) and fast retina keypoint (FREAK) (Alahi et al., 2012).

In the former comparative study, the authors found that the GLOH descriptor gets the best results in most of the tests they conducted, it was closely followed by SIFT. The same authors stated that those results highlight the robustness and distinctive character of region-based SIFT descriptors. In the latter comparative analysis, it has been concluded that the SIFT and SURF descriptors, followed by ORB, give the best results when considering scale and orientation changes. Thus, they adopted SIFT in their application.

Hence, the best choice for our application would be either SIFT or SURF. We opted for the latter for it needs lesser calculation time than that needed by SIFT (Fu, 2014), as said above; the devised system is to be implemented in a real-time application. The pictures given in Figure 2 show the SURF matching of different frames to one particular image, called *key picture* as in (Shoaib, 2009) in the rest of the paper.

It is worth to mention that SURF technique works on greyscale images and finds pair-by-pair matches between the key-picture primitives and those of the frame being matched. Only good matches are kept, which allows evaluating the homography connecting primitives to their projections. Such homography is found through the use of the Random Sample Consensus (RANSAC) algorithm, a similar use of homography matrix in Zhang et al. (2012) and Wu and Fang (2007) proved its efficiency. The images presented in Figure 2 illustrate the robustness of SURF to photometric change effect [Figure 2(a)], orientation change [Figure 2(b)], scale change [Figure 2(c)] and occlusion [Figure 2(d)]. The frame size is set to 640×480 pixels, except for the second figure (corresponding to scale change, set to 320×240 pixels). As shown in the figure, the SURF descriptor shows a high robustness towards geometric and photometric changes.

2.2 Perspective projection

From a mathematical point of view, finding the position and orientation of the 2D object projection on a plane is equivalent to finding the 2D-2D perspective projection matching, described by a planar homography.

In fact, if $[x \ y \ 1]$ represents the homogeneous coordinates of a point X of the object and $[x' \ y' \ 1]$ represents the homogeneous coordinates of its projection X' on the plan, then X is related to X' by the equation: $X = H X'$ with H , given in formula (1), is a 3×3 matrix representing a planar homography. Every match generates two linear equations, equations (2) and (3).

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (1)$$

$$x'(h_{31}x + h_{32}y + h_{33}) = h_{11}x + h_{12}y + h_{13} \quad (2)$$

$$y'(h_{31}x + h_{32}y + h_{33}) = h_{21}x + h_{22}y + h_{23} \quad (3)$$

So, in order to find the matrix H , which is an 8-degree of freedom matrix, we need four matches of four non-collinear points to generate the eight linearly-independent equations needed to solve the system. The RANSAC algorithm is exploited to select four correct matches from all those generated by the matching process.

Figure 2 SURF matching robustness to (a) photometric change, (b) orientation change, (c) scale change and (d) partial occlusion (see online version for colours)

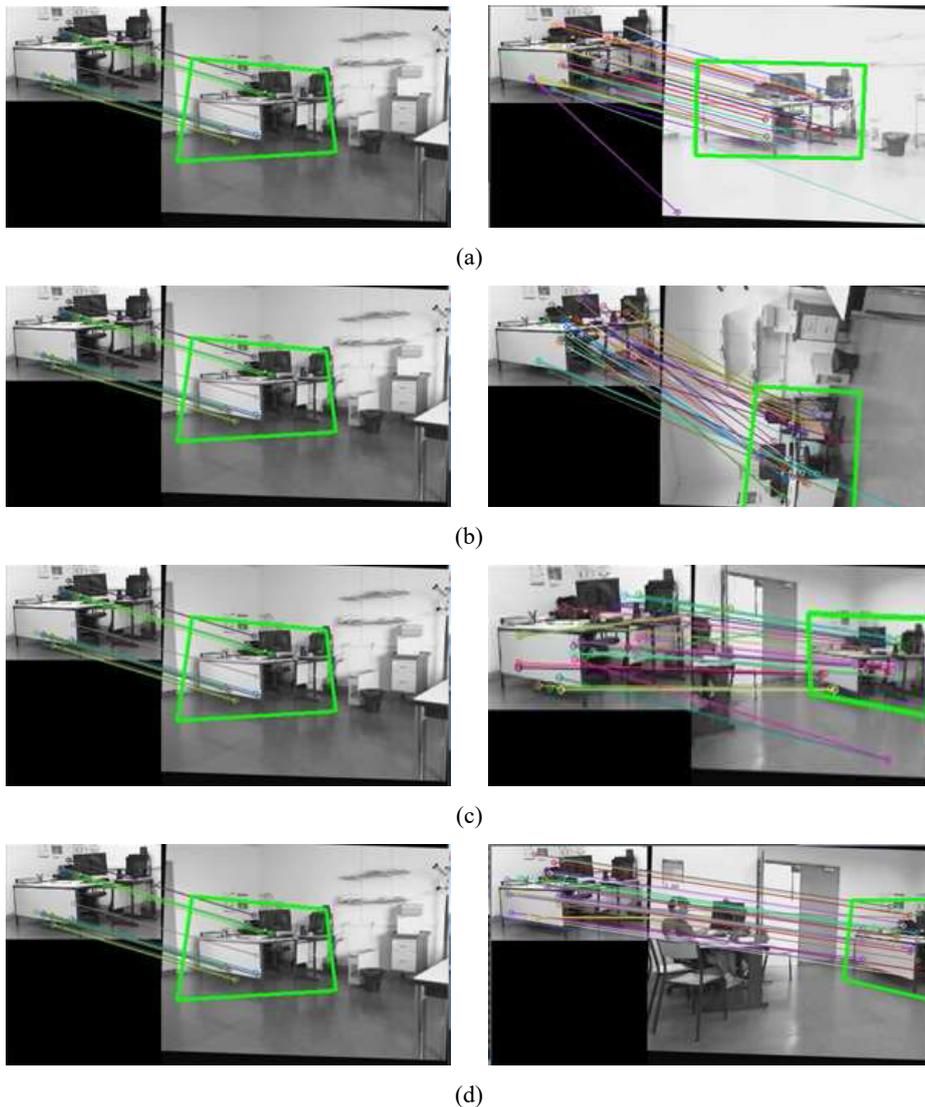
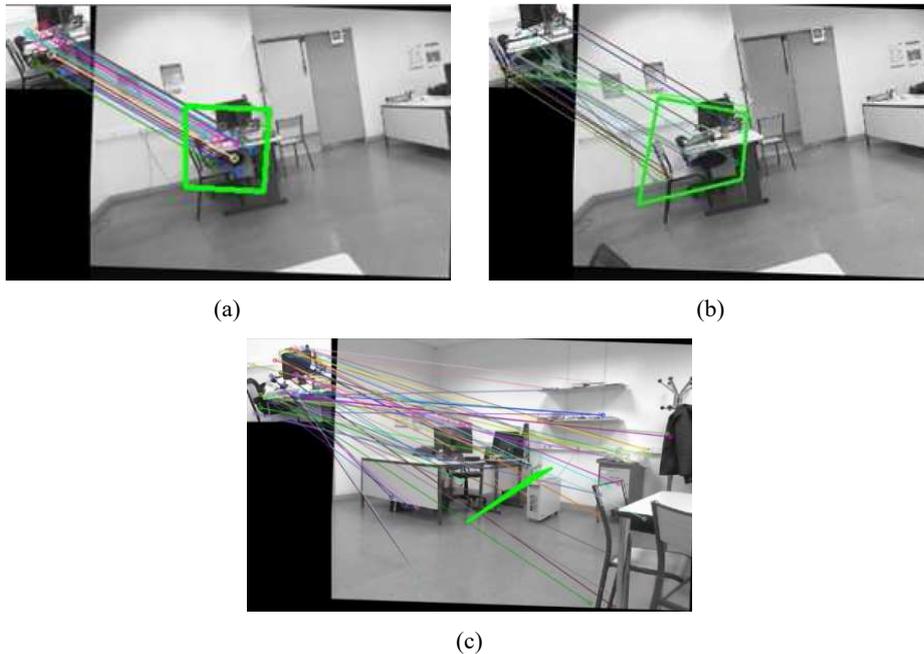


Figure 3(a) shows the result of an image perspective projection using the RANSAC algorithm after matching by SURF. Figure 3(b) shows a partial deformation of the perspective projection, caused by the movement of an object (chair) in the scene. Finally,

Figure 3(c) shows a bad matching resulting in a severely distorted perspective projection. To achieve a better systematicity of the proposed algorithm, we put the following rule. Since the object being matched is a square or rectangular image, its projection must be a convex quadrilateral. So, the matching is judged to be bad if it gives less than four matches, or if the resulting projection is a concave quadrilateral.

Figure 3 Perspective projection of matched images, (a) a good matching (b) a slight deformation of the projection due to an object move on the scene (c) a bad matching resulting in a severe deformation of the perspective projection (see online version for colours)



3 The proposed approach

In this section, details about the design and motivation of choices made in the proposed algorithm are provided.

3.1 Motivation

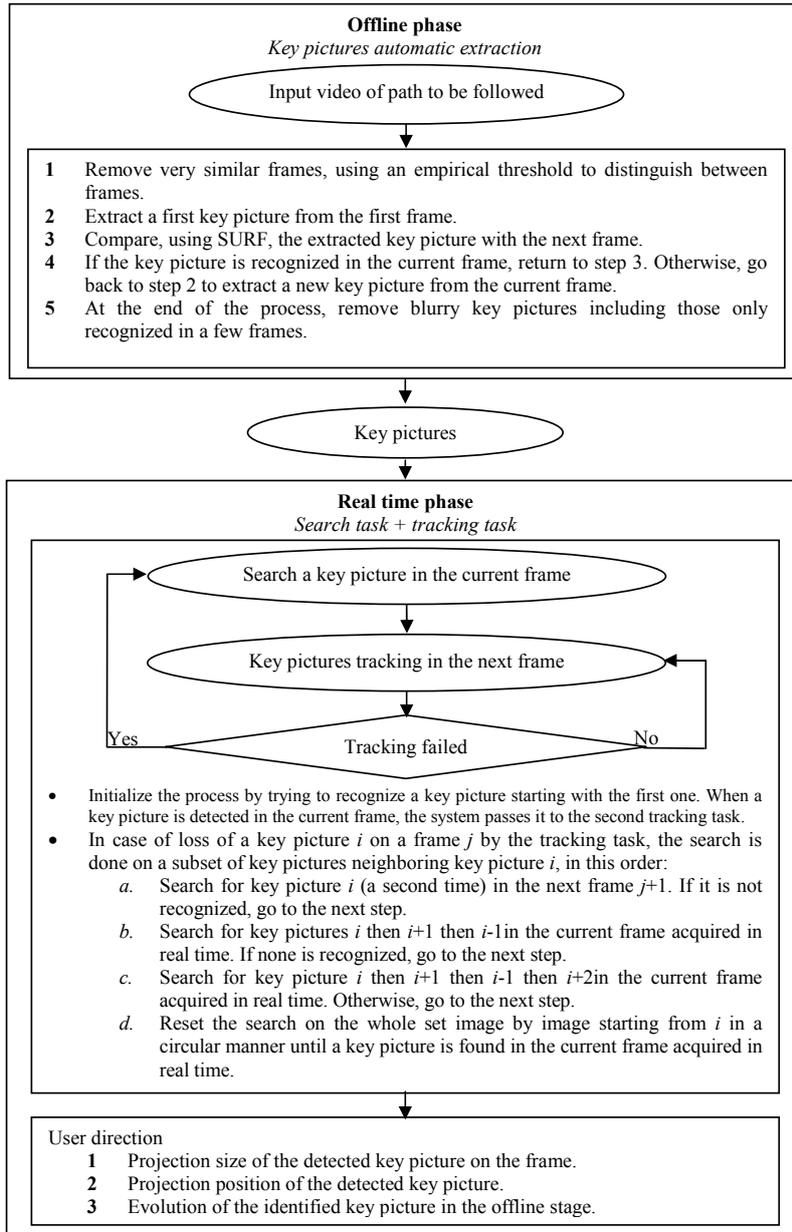
The main motivation behind the present work is to provide users (APs) with an automatic system that does not require any user intervention when selecting key pictures. Taking as example the system we used our approach in, indoor AP directing; in such a system, it would be impossible for the patient to choose he himself the key pictures to be considered in when matching images; and even his caregiver would have difficulty to manually accomplish such a task.

Hence, the originality of the present work is not in the tracking task itself; the novelty here is in the automatic choice of key pictures to help in the tracking operation. Actually,

there is no previous work that assures the automatic selection of key pictures. This is the plus of the applicability of the proposed approach on the field of medicine.

Our method, thus, differs from previous works in two aspects; on the one hand, the selection of features in our proposal is automatic and does not require any user intervention, and on the other hand, it is mainly oriented to APs, it helps them by showing them directions on the augmented reality glasses.

Figure 4 The two stages of the proposed method



3.2 The proposed algorithm

The proposed approach mainly consists of two stages: key pictures automatic extraction and the search and tracking tasks (see Figure 4); this architecture is inspired from (Elloumi et al., 2013). A first stage, which is offline, allows the extraction of small key pictures, also called primitives, from a video of the path to be followed (see Section 3.2.1). The second stage, the real-time stage (see Section 3.2.2), allows monitoring the path on the basis of the key pictures extracted in the first stage. The latter stage is itself composed of two intermittent tasks: the search task and the key picture tracking task.

3.2.1 Key picture automatic extraction

The first stage of the proposed method was developed to automatically select a set of natural features, called here key pictures, which will be the basis of the actual path tracking presented in the second stage. The efficiency of the tracking mostly depends on the quality of the first stage results.

The automatic nature of this stage is essential; otherwise a manual extraction of key pictures would need to be considered, which can almost be restricted to an expert who knows how to select the right key pictures in the right frames. Obviously, the patient cannot accomplish this task; even the caregiver would be unable to do it. Thus, the purpose of the first stage corresponds to the automatic selection of a minimal number of key pictures that have the best chance to be recognised during the second stage. The input of the first stage is a video of the path to be followed, while the output is a set of key pictures that will be used to locate the user in the second stage.

We put and respected three constraints regarding key picture automatic extraction:

Rule 1 The selection of images that best fit with recognition.

Rule 2 The proper choice of the key pictures size.

Rule 3 The selection of the minimum number of key pictures to reduce the search time during the second stage as much as possible.

For what concerns rule 1, images that best fit the recognition are those with a large number of detected primitives, since primitive detection is the foundation of the second stage recognition, this is detailed in the next section. It is worth to be mentioned here that the primitives extracted and exploited at this point are basic primitives (offered by SURF), not the key pictures.

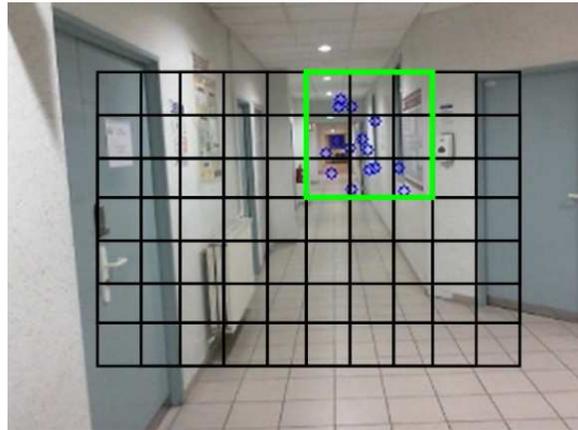
As far as rule 2 is concerned, the average size of images which has given the best recognition results is 250×250 pixels, however to obtain optimal results, the size choice was automatically selected for every single path (see the experimentation part).

Finally, for rule 3, to perform a selection of a minimal number of images, we have to choose the key pictures that have the best chance to be recognised in a lot of frames, over a relatively long path. This was interpreted in the present method by a frame extraction from the central part of the frame at hand; the picture parts that are on edges may not be visible when moving; the extraction is only done on the central area of the frame. Similarly, blurred images recorded during motion are rejected. Thus, a second check is performed to remove all non-relevant key pictures that are only recognised in a few frames.

To extract a key picture on a given frame, we scan using a mask window along the central part of the frame and we record the detected primitive count, then we select the area which has the highest count of found primitives (Figure 5).

There is no need to extract another key picture as long as the current one is identified in the next frames. Once the key picture is lost, a new one is extracted.

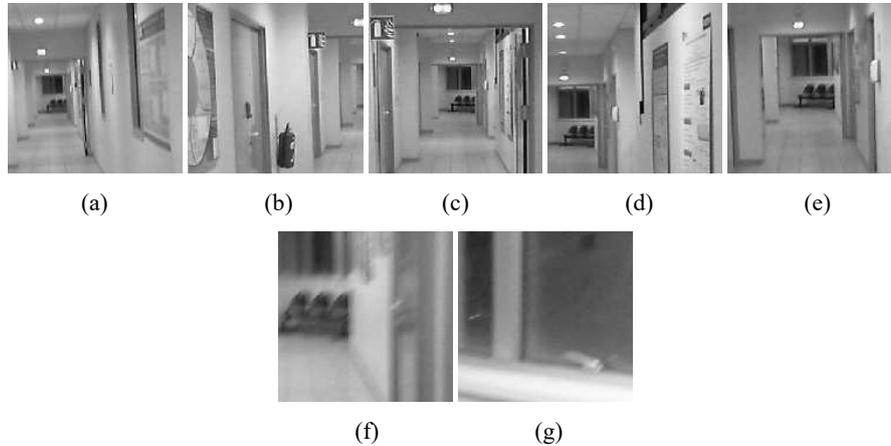
Figure 5 Key pictures automatic selection (green) after scanning black areas (see online version for colours)



Hence, on the whole, the first stage proceeds as follows:

- 1 Remove very similar frames without noticeable changes in the scene. This is technically realised through an image differentiation: the next frame is only taken into account if its difference with the current one is higher than an empirically chosen threshold.
- 2 Extract a first key picture from the first frame.
- 3 Compare the obtained key picture with the next frame. The comparison here is made within the SURF method.
- 4 If the key picture is recognised in the current frame, we return to Step 3. Otherwise, we go back to Step 2 to extract a new key picture from the current frame, and so on.
- 5 At the end of the process, we realise a second scan to detect and remove blurry key pictures including those only recognised in a few frames [the last two key pictures in Figures 6(f)–6(g)]. And finally, in this way, we will have our set of key pictures, ready to be used in the second stage [the five key pictures in Figure 6(a)–6(e)].

During the whole key-picture extraction process, a log file of key picture evolution is created, to keep their projection sizes and positions in every frame where they have been recognised. Those parameters will be used in the second stage to compute the arrow orientation.

Figure 6 The set of obtained key pictures before removal of blurred images [Figure 6(f)–6(g)]

3.2.2 Real time path tracking

The purpose of this stage is to spot and locate the key pictures on the path to be followed. Two tasks are intermittently executed:

- Task 1 Seek the possible presence of one of the key pictures in the current frame. This search is done by matching the SURF results between the current frame and the key pictures, according to an order which will be explained in the next section.
- Task 2 Lucas-Kanade optical flow in pyramids (Shi and Tomasi, 1994) is then employed to track the identified key picture in real time. In case of target loss by the optical flow, we go back to the first task (Figure 4, real-time phase).

Since the first task goal is to find one of the key pictures to give it to the second tracking task, what is needed is a fast search policy. To do so, all key pictures are extracted and sorted in their recognition order. To reduce the computational time, the search is restricted to a subset of key pictures neighbouring the last recognised key picture. The size of such a set proportionally depends on the time elapsed from the last recognised key picture. This has been achieved by the following actions.

- Initialise the process by trying to recognise a key picture starting with the first one. When a key picture is detected in the current frame, the system passes it to the second tracking task.
- In case of loss of a key picture i on a frame j by the tracking task, the search is done on a subset of key pictures neighbouring the key picture i , in this order:
 - a Search for key picture i (a second time) in the next frame $j + 1$. If it is not recognised, go to the next step.
 - b Search for key pictures i then $i + 1$ then $i - 1$ in the current frame acquired in real time. If none is recognised, go to the next step.

- c Search for key picture i then $i + 1$ then $i - 1$ then $i + 2$ in the current frame acquired in real time. Otherwise, go to the next step.
- d Reset the search on the whole set image by image starting from i in a circular manner until a key picture is found in the current frame acquired in real time.

Each one of the previous steps is achieved on the current frame. Therefore, in case of failure of recognition and of execution of Steps a, b and c, we will not be on frame number $j + 3$ but much further away, since SURF could not cope with the video acquisition speed (Bay et al., 2008), even if SURF is less complex than SIFT, it remains slower than the video speed. This gives enough time to decide whether the search was successful or not, in the latter case, we go to Step d, which is the circular search of the whole set from image i , until finding a key picture.

3.2.2.1 User direction

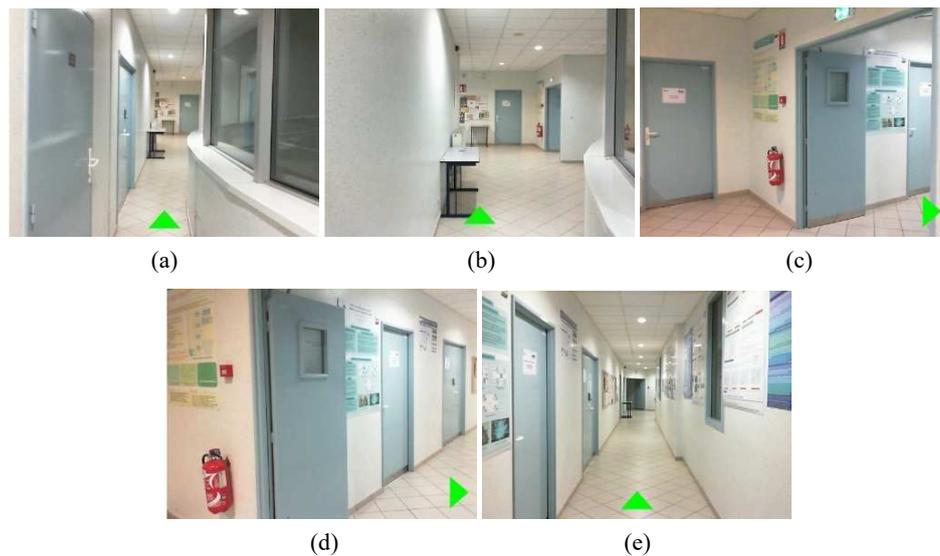
Once the second stage has located a key picture, the orientation of the arrow to display can be deduced on the basis of the three following points:

- 1 projection size of the detected key picture on the frame
- 2 projection position of the detected key picture
- 3 evolution of the identified key picture in the offline stage.

Actually, a given key picture can be detected on a long path, and from the observer point of view, the key picture projection size changes according to his position. Inversely, knowing the key picture projection, position and size, the approximate position of the observer on the path can be deduced. His direction is then possible regarding the history of key picture evolution in the offline stage.

In this history, positions and sizes of key-picture projection in every frame are correctly saved. Figure 7 shows a user direction example.

Figure 7 User direction example, (a) frame no. 123 (b) frame no. 245 (c) frame no. 512 (d) frame no. 576 (e) frame no. 645 (see online version for colours)



A video clip illustrating the second stage of the proposed method is provided as supplementary material: <http://youtu.be/uT-wndf9Irc>.

4 Results and discussion

To test the performances of the proposed method, we applied it on a set of six different indoor paths from 20 to 30 meters of length. The platform used for the tests was the C++ version of the OpenCV library. The obtained results are analysed in the following paragraphs.

The number of key pictures found varied from 4 to 13 depending on the path to be tracked. The graphs presented in Figure 8 correspond to the second real-time stage, on one path under different photometric conditions. The first graph was realised under common photometric conditions, while the second is the result of a case with brightness deterioration. In both graphs, the x axis presents the frame number, while the y axis represents the key picture index. In the graphs, the blue line continuity presents a normal tracking implemented by the optical flow (Shi and Tomasi, 1994) in this work, while the red line presents a search process. In simple words, if the key picture is kept, the tracking phase carries on; otherwise a new key picture is searched.

The First graph, presented in Figure 8(a), corresponds to a good global tracking, where transitions between targets (key pictures) are all in the direction of the path progression, and every key picture extracted by the first stage has been used in the real-time tracking phase.

Since tracking was good, we find that, as seen from the same graph, needed transitions (marked in red on the graph) were really little. The execution was made in real time and the slowdown, caused by the search task, was almost unnoticeable. The SURF method is algorithmically complex when taking into account real-time contexts. For this reason, we opted for the use of the optical flow, since it ensures a real-time execution of the method.

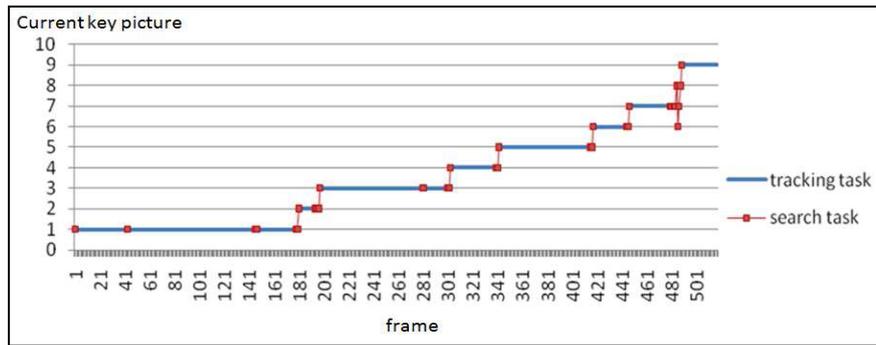
The second graph, presented in Figure 8(b), illustrates the case of a weak global tracking; which was caused by the large number of matching attempts. In fact, a large number of matching failures, graphically illustrated by a large number of fluctuations marked in red, indicates that the search task was unable to precisely locate any recorded key pictures, which resulted in an expensive calculation time. This number of fluctuations that participates in the computation of what we call *failure rate*, is used hereafter as a comparison and performance metric.

For the two graphs on Figure 8, the failure rate was 4% for Figure 8(a) and 22% for Figure 8(b). This rate was computed by dividing the number of erroneous frame matching by the total number of frames. A frame matching is considered erroneous if the system is unable to locate or track a particular key picture.

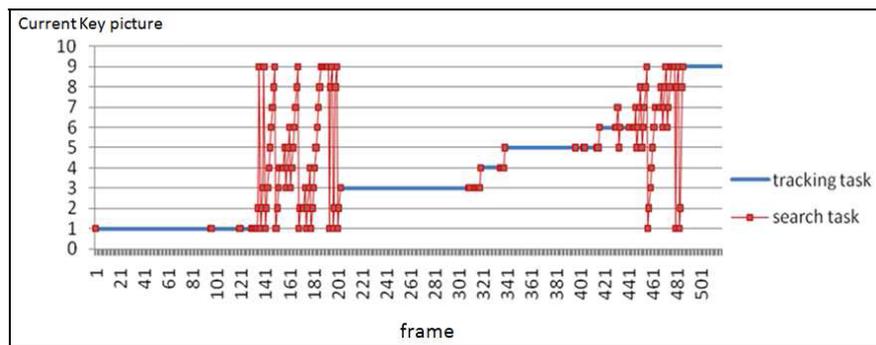
The size of the mask window used in scanning the central part of a frame to extract a key picture, also the size of the latter, is automatically defined in the offline phase. Herein, the system performs the extraction of key pictures according to a given resolution, it calculates the failure rate on the same input path using these key pictures, and repeats the same process for other resolutions till finding the least failure rate. Eventually, the resolution that gives the minimum failure rate is retained in addition to its

corresponding key pictures. Consequently, the process allows getting optimum results and associates to each path the fittest key picture resolution.

Figure 8 Representation of the search and tracking tasks of the second phase in an (a) original path and (b) after brightness deterioration of the same environment (see online version for colours)

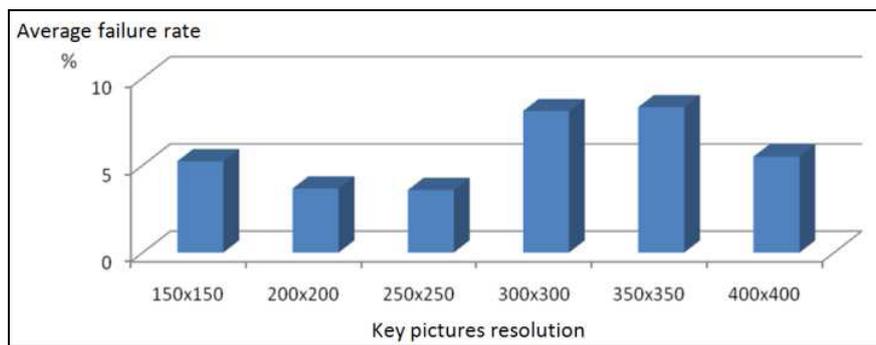


(a)



(b)

Figure 9 Average failure rate according to key pictures resolution (see online version for colours)



The graph given in Figure 9 shows that a 250×250 pixel resolution gives the least average failure rate over the six tested paths, but the typical standard deviation illustrated in Figure 10 presents large values and indicates that this average resolution cannot therefore be adopted for all paths.

Figure 10 Standard deviations of failure rate with respect to different key picture resolutions (see online version for colours)

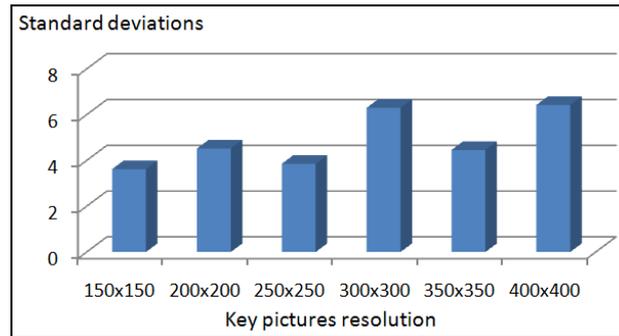


Figure 11 Frame photometric degradation, (a) original frame (b) contrast reduction ($\alpha = 0.4$) (c) brightness reduction ($\beta = -100$) (d) addition of Gaussian noise (mean = -40 , variance = 35) (see online version for colours)



(a)



(b)



(c)



(d)

In a similar way, to test the robustness of the method towards photometric changes, we have modified the contrast and brightness of the six benchmark paths and calculated the failure rate for every situation.

The photometric change is formulated in equation (4), where $f(x, y)$ represents the intensity of a pixel in the original colour image, $g(x, y)$ corresponds to the resulting pixel intensity, α represents the gain which controls the contrast and β represents the angle which controls brightness.

$$g(x, y) = \alpha \cdot f(x, y) + \beta \quad (4)$$

We ran the three following photometric changes (see Figure 11 of a typical example):

- contrast reduction [Figure 11(b)] by modifying α in equation (4)
- brightness reduction [Figure 11(c)] by modifying β in equation (4)
- adding random Gaussian noise [Figure 11(d)].

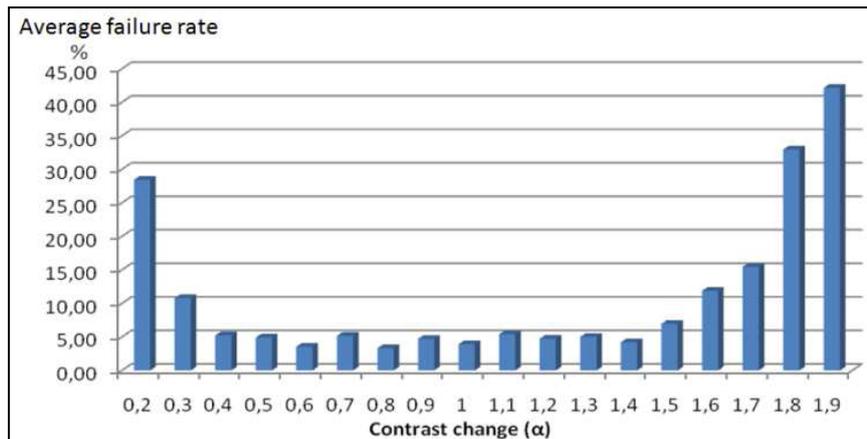
For example, adding a Gaussian noise with a mean fixed to -40 and a variance set to 35 is equivalent to a change in the colour intensities of each pixel by adding a random value in the range $[-40 - 35, -40 + 35]$.

The two graphs given in Figure 12 display the average failure rate according to contrast change [Figure 12(a)] and brightness change [Figure 12(b)], respectively, over the six benchmark paths.

According to the last two graphs, modifying the α values within the range $[0.4, 1.5]$, or the β values within the range $[-70, +80]$ does not degrade the performances of the method. The average failure rate over the total number of frames is lower than 7%.

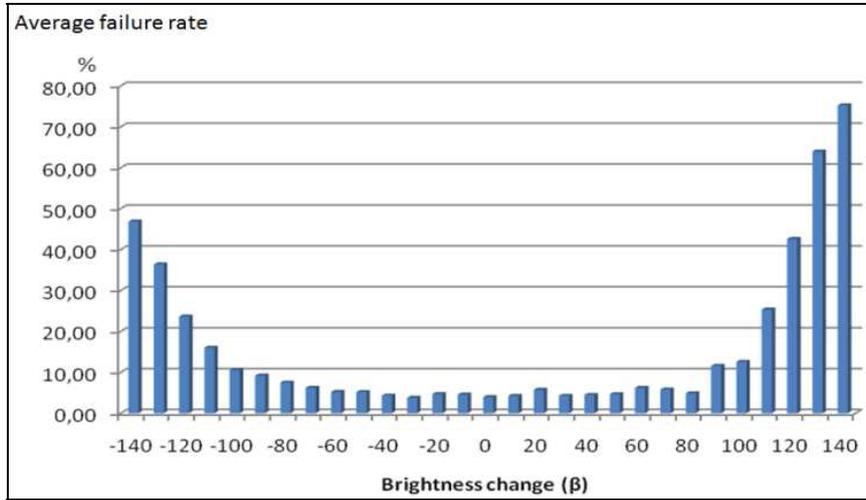
Similarly, the graph of Figure 13 illustrates the performances of the method after the addition of some artificial noise, generated here by a Gaussian noise with variable mean and variance. This graph displays the average failure rate for all the six paths.

Figure 12 Average failure rate with (a) contrast change and (b) brightness change (see online version for colours)



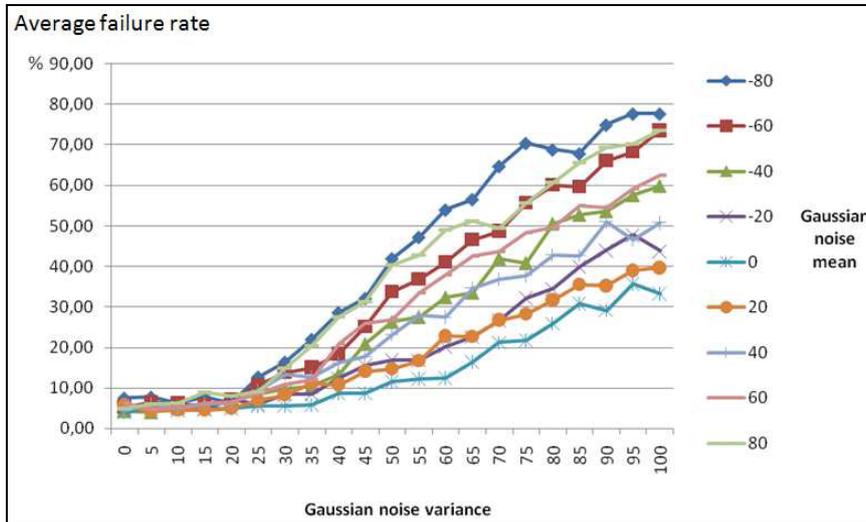
(a)

Figure 12 Average failure rate with (a) contrast change and (b) brightness change (continued) (see online version for colours)



(b)

Figure 13 Average failure rate in the presence of variable variance Gaussian noise (see online version for colours)



A Gaussian noise with a variance located in the [0, +20] range corresponds to a failure rate lower than 7%. Also, the addition of a Gaussian noise with a variance lesser than 25 gives a failure rate lower than 10%.

On the whole, the obtained results highlight the robustness of the proposed method with regard to photometric variation.

Four video clips illustrating examples of successful path tracking are provided as supplementary material on the links:

- http://youtu.be/o_Q7pWORx_4 (original photometric conditions)
- <http://youtu.be/2JcM7JNC6xI> (contrast reduction $\alpha = 0.4$)
- <http://youtu.be/mbRJe7EyBrk> (brightness reduction $\beta = -100$)
- <http://youtu.be/0Ua1e8BRQ4k> (Gaussian noise mean = 0.4, variance = 35).

These results globally highlight the robustness of the proposed method in regards to photometric variations.

5 Conclusions

In this work, we proposed a visual method that automatically selects and tracks natural features in an indoor environment. The system is mainly dedicated to assist APs in their indoor moves.

This method, which will be used for the development of an application on augmented reality glasses, does not require any user intervention, it helps them by showing them directions on the augmented reality glasses.

The method is mainly composed of two stages: the first one aims at selecting a set of relevant key pictures from the source path video. The second stage consists of using these key pictures for path recognition and for directing the patient along the path in real time. Using this information, the system computes and displays direction arrows that are employed to orientate and guide the patient in his movements.

The recognition is realised by matching the SURF descriptor, and the real-time constraint compliance is achieved, on the one hand, by combining the matching with the Lucas-Kanade in pyramids optical flow tracking process, and, on the other hand, by searching through the closest key pictures when the currently used one is lost.

The experimental results show that the proposed method is very promising and gives satisfactory results in a static and well-contrasted environment, and that is robust enough to photometric changes.

There are some limitations of this work that should be pointed out when considering the results presented in this paper. It should be noted that the quality of the tracking depends on the characteristics of the undertaken path. The method works properly in static environments, rich of contrast and texture, but it is less adapted to dynamic or low-contrasted environments.

Actually, objects and parts of the scene may move in a dynamic environment. The features and key pictures that are extracted by the proposed method can be occluded by other objects and may totally disappear thus jeopardising the recognition process. Similarly, the system cannot detect primitives when considering low contrasted environments.

As a future direction of research, we intend to improve the proposed method by automatically adjusting frame contrast. The method can be further enhanced by automatically updating the key pictures, to support more constraints about the working environment dynamics. An intuitive way to do this is associating weights to several key

pictures selected in the same frame, and then by updating these weights during the tracking phase.

Acknowledgements

The authors would like to thank Dr. Amer Draa for the help he provided when writing this paper.

References

- Alahi, A., Ortiz, R. and Vandergheynst, P. (2012) 'FREAK: fast retina keypoint', *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.510–517.
- American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders, DSM-5®*, American Psychiatric Pub., Arlington VA.
- Bai, Y., Jia, W., Zhang, H., Mao, Z.H. and Sun, M. (2014) 'Landmark-based indoor positioning for visually impaired individuals', *2014 12th International Conference on Signal Processing (ICSP)*, pp.668–671.
- Bay, H., Ess, A., Tuytelaars, T. and van Gool, L. (2008) 'Speeded-up robust features (SURF)', *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp.346–359.
- Belongie, S., Malik, J. and Puzicha, J. (2002) 'Shape matching and object recognition using shape contexts', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp.509–522.
- Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C. and Fua, P. (2012) 'BRIEF: computing a local binary descriptor very fast', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 7, pp.1281–1298.
- Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G. et al. (2007) 'Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria', *Lancet Neurol.*, Vol. 6, No. 8, pp.734–746.
- Elloumi, W., Guissous, K., Chetouani, A., Canals, R., Leconge, R., Emile, B. and Treuillet, S. (2013) 'Indoor navigation assistance with a smartphone camera based on vanishing points', *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp.1–9.
- Freeman, W.T. and Adelson, E.H. (1991) 'The design and use of steerable filters', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 9, pp.891–906.
- Fu, W. (2014) *Visual Servoing for Mobile Robots with Collision Avoidance and Field-of-View Constraints*, PhD thesis, University of Evry, Evry, France.
- Hile, H. and Borriello, G. (2008) 'Positioning and orientation in indoor environments using camera phones', *IEEE Computer Graphics and Applications*, Vol. 28, No. 4, pp.32–39.
- Ke, Y. and Sukthankar, R. (2004) 'PCA-SIFT: a more distinctive representation for local image descriptors', *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.506–513, IEEE Computer Society, Washington, DC, USA.
- Koenderink, J.J. and van Doorn, A.J. (1987) 'Representation of local geometry in the visual system', *Biol. Cybern.*, Vol. 55, No. 6, pp.367–375.
- Lazebnik, S., Schmid, C. and Ponce, J. (2003) 'A sparse texture representation using affine-invariant regions', *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings*, Vol. 2, pp.II-319–II-324.
- Leutenegger, S., Chli, M. and Siegwart, R.Y. (2011) 'BRISK: binary robust invariant scalable keypoints', *2011 IEEE International Conference on Computer Vision (ICCV)*, pp.2548–2555.

- Lowe, D.G. (1999) 'Object recognition from local scale-invariant features', *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, pp.1150–1157.
- Matas, J., Chum, O., Urban, M. and Pajdla, T. (2004) 'Robust wide baseline stereo from maximally stable extremal regions', *Image and Vision Computing*, Vol. 22, No. 10, pp.761–767.
- Mikolajczyk, K. and Schmid, C. (2005) 'A performance evaluation of local descriptors', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp.1615–1630.
- Nozawa, M., Hagiwara, Y. and Choi, Y. (2012) 'Indoor human navigation system on smartphones using view-based navigation', *2012 12th International Conference on Control, Automation and Systems (ICCAS)*, pp.1916–1919.
- Quintana, E. and Favela, J. (2013) 'Augmented reality annotations to assist persons with Alzheimers and their caregivers', *Pers Ubiquit Comput*, Vol. 17, No. 6, pp.1105–1116.
- Reisberg, B. and Franssen, E.H. (1999) 'Clinical stages of Alzheimer's disease', *An Atlas of Alzheimer's Disease*, Taylor & Francis, New York.
- Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011) 'ORB: an efficient alternative to SIFT or SURF', *2011 IEEE International Conference on Computer Vision (ICCV)*, pp.2564–2571.
- Schaffalitzky, F. and Zisserman, A. (2002) 'Multi-view matching for unordered image sets, or 'how do I organize my holiday snaps?''', *Proceedings of the 7th European Conference on Computer Vision – Part I*, pp.414–431, Springer-Verlag, London, UK.
- Shi, J. and Tomasi, C. (1994) 'Good features to track', *1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94*, pp.593–600.
- Shoib, M. (2009) 'Key picture error concealment using residual motion-copy in scalable video coding', *2009 WRI World Congress on Computer Science and Information Engineering*, pp.591–595.
- Thomas, P., Billon, R., Hazif-Thomas, C., Lalloue, F., Pariel, S., Inscale, R., Belmin, J. and Clement, J.P. (2006) 'Pronostic of institutionalisation for demented out patients. The PIXEL study', *Geriatrics Journal*, Vol. 31, No. 8, pp.589–596.
- Tuytelaars, T. and Mikolajczyk, K. (2008) 'Local invariant feature detectors: a survey', *Found. Trends. Comput. Graph. Vis.*, Vol. 3, No. 3, pp.177–280.
- World Health Organization and Alzheimer's Disease International (2012) *Dementia: A Public Health Priority*, pp.7–8, Geneva, Switzerland.
- Wu, F. and Fang, X. (2007) 'An improved RANSAC homography algorithm for feature based image mosaic', *Proceedings of the 7th WSEAS International Conference on Signal Processing, Computational Geometry & Artificial Vision*, pp.202–207, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA.
- Zhang, M., Hou, Y. and Hu, Z. (2012) 'Accurate object tracking based on homography matrix', *2012 International Conference on Computer Science Service System (CSSS)*, pp.2310–2312.