2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Scale space and free space topology analysis for omnidirectional images

Romain Marie, Ouiddad Labbani-Igbida and El Mustapha Mouaddib

Abstract— **This work is part of a global framework that aims to build a complete set of algorithms for autonomous robot exploration, navigation and map building using a sole omnidirectional camera. It focuses on topology extraction of navigable free spaces to yield autonomous robot exploration. The paper proposes a new algorithm, the Omnidirectional Delta Medial Axis (ODMA), with linear-time implementation to extract robust topology of the robot's local free space. It first introduces an adapted metric to cope with the deformations involved by the catadioptric sensor, and that is conformal with the ground 2D-metric. Then, it produces a pruned skeleton of the free space, inspired by the delta medial axis. Experimental results validate the approach for both skeleton precision and stability with respect to the robot position and free space topology.**

Index Terms— **Omnidirectional vision, Local topology extraction, robot safe navigation.**

I. Introduction and related work

To be fully autonomous, a mobile robot needs a safe and efficient solution to navigate in a scene. When the only available sensor is a camera, it becomes a very challenging task, especially when the environment is initially totally unknown. Promising results have been recently obtained with appearance based approaches. Typically, existing methods ([13], [14], [4]) proceed in two steps. First, the robot is manually driven through the environment (learning or training step), and builds a succession of representations containing all the information required to replay the trajectory. Then, in a second time, the robot navigates autonomously on the previously driven path, by comparing its current perception to the memory he built earlier.

A structure from motion algorithm is used in [13] to build a 3D map of the explored area based on a reference video sequence. In the navigation step, the robot uses this map to locate itself and follows the previous trajectory. In work [14], both key-images and interest points are constructed during the teaching step and then used by the robot as a hybrid topological-geometrical representation of the environment. The robot navigates by moving from a key image to the next one and tracking the expected position of each mapped landmark. In [4], once again landmarks are detected and tracked. The replay step involves comparing the coordinates of each landmark with the ones acquired in the training step and then moves the robot to minimize the difference. All those solutions involve that the robot needs to follow almost exactly the same path as in the learning phase, since when it deviates, the constructed visual features might fail to be recognized.

Only a few work consider concurrent exploration and navigation based on the local perception and discovery of the free space. The robot autonomously moves by tracking different visual cues that define the navigable space. In [11], edge extraction and analysis is performed in order to find lane structures such as boundaries, crossings or stop lines. The authors, in [12], use a combination of visual cues and ladar-derived structural information for segmenting trails for autonomous navigation. The proposed shape-based visual trail tracker assumes that the approaching trail region is approximately triangular under a perspective camera; a contrast measure is then used to classify the expected trail regions. In [2], the authors use a k-means approach to perform a fast texture segmentation on outdoor perspective images. A natural path extraction algorithm is then applied for autonomous navigation purposes.

The work presented in this paper is based on [10], where an active contour algorithm is proposed to extract in real time the boundaries of the navigable space using an omnidirectional image. The latter also includes consideration to the false obstacles (white lines on the road for instance) that could stand inside this portion of the image. However, the boundaries of the ground are hardly a sufficient control input to insure a structured exploration of the environment in the safest possible way. In [15], the authors suggest an efficient navigation methodology based on the instantaneous local Voronoi diagram extracted from a laser range finder. To produce a similar result using instead an omnidirectional sensor, it is required to extract this diagram directly in the image.

The main contribution of this paper is to propose a new solution for shape topology extraction, specifically designed for ground analysis in omnidirectional images. It considers the distortions involved by the catadioptric sensor and the specific position of the camera with respect to the ground, to produce a corrected distance directly in the image. From this specific metric, this paper proposes an adapted version of the δ medial axis [9], a fast and efficient solution for pruned skeleton extraction, to model the local 3D ground's Voronoi diagram directly in the image.

The paper is organized as follows: In Section II, we present how to extract shape topology using medial axis techniques. In Section III, the geometry of the catadioptric sensor is used to propose an adapted omni-metric and an extended version of the δ-medial axis algorithm, in order to model the topology of the free space directly in the image. Section

*This work was supported by ANR agency under RDiscover project, and financed by DGA (Direction Générale de l'Armement) and FEDER (Fond Europeen de DEveloppement Regional) in Picardie.

The authors are with Modélisation, Information & Systèmes lab., University of Picardie Jules Verne, Amiens, France firstname.lastname@u-picardie.fr

IV presents experimental results comparing the standard and adapted version of the δ-medial axis and proving the necessity of the adapted model.

II. Medial axis and shape topology extraction

A. General Medial Axis definition

Let X be a subset of $\mathbb{R}^2$ and ∂X its boundary. Initially defined by Blum [3], the real medial axis $S_{\mathbb{R}^2}$ of X (also named the skeleton of X) is the set of points $\mathbf{x} \in X$ that are centers of maximal disks fully included in X but not in any other disk included in X.

Let us introduce the feature transform $FT(\mathbf{x})$, corresponding to the projection of $\mathbf{x}$ on ∂X, i.e. the set of points in ∂X that minimize the distance from $\mathbf{x}$ to the boundary ∂X:

$$FT(\mathbf{x}) = \{\mathbf{y} \in \partial X | \forall z \in \partial X, d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{x},z)\} \quad (1)$$

Note that the feature transform depends on a metric projection to determine the nearest point mapping. The euclidian distance $d(\mathbf{x},\mathbf{y}) = d_E(\mathbf{x},\mathbf{y}) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$ is usually considered, and the euclidian feature transform produced is indifferently denoted $FT(\mathbf{x})$ or $FT_E(\mathbf{x})$.

Clearly, the centers of Blum's maximal disks have at least two closest points in ∂X. The skeleton $S_{\mathbb{R}^2}$ can consequently be expressed in term of $FT(\mathbf{x})$:

$$S_{\mathbb{R}^2} = \{\mathbf{x} \in X,\ |FT(\mathbf{x})| \geq 2\} \quad (2)$$

with $|FT(\mathbf{x})|$ the cardinal of $FT(\mathbf{x})$.

Transposing directly this definition to discrete subspaces of $\mathbb{Z}^2$ is however problematic, since in most cases, $|FT(\mathbf{x})| = 1$, even when $\mathbf{x}$ is actually the discrete approximation of a point belonging to the object's exact skeleton. Let us for instance consider the case of an horizontal ribbon defined in $\mathbb{Z}^2$, with constant even width and infinite length. Clearly, using the previous definition, we have $S_{\mathbb{Z}^2} = \emptyset$, while the corresponding real medial axis $S_{\mathbb{R}^2}$ is the locus of the points composing the centre line of the ribbon.

Therefore, the definition of the discrete skeleton needs to be slightly modified. We first consider the reduced feature transform $ft(\mathbf{x})$ [7] corresponding to the first element of $FT(\mathbf{x})$ with respect to a lexical ordering. Then, given $X \subset \mathbb{Z}^2$, Hesselink et Roerdink [7] proposed to define the skeleton $S_{\mathbb{Z}^2}$ of X as composed by the set of points $\mathbf{x} \in X$ that verify:

$$\exists \mathbf{y} \in X, ||\mathbf{x}-\mathbf{y}|| = 1 \ s.t. \begin{cases} ||ft(\mathbf{x}) - ft(\mathbf{y})|| > 1 \\ \& \\ ||\mathbf{m} - ft(\mathbf{y})|| \leq ||\mathbf{m} - ft(\mathbf{x})|| \end{cases} \quad (3)$$

where $\mathbf{m}$ is the midpoint of the line segment $[\mathbf{x},\mathbf{y}]$.

The first condition considers the (euclidian) distance between the feature transforms of $\mathbf{x}$ and $\mathbf{y}$. The threshold is set strictly greater than 1, so that skeleton points due to the discreteness of the background are ignored. The second condition makes the skeleton as thin as possible since it insures that only the closest point to the real skeleton position is added to $S_{\mathbb{Z}^2}$. Note that if $\mathbf{x}$ and $\mathbf{y}$ are equally close, both are included.

B. Medial axis pruning

The previous formulation (Eq. (3)) leads to a good approximation of the real medial axis after smoothing of the boundary, but also yields many spurious branches (Fig. 1) due to small contour perturbations. The suppression (pruning) of such undesired effects is necessary for image analysis and is often used in some postprocessing phase.

This work is based on the δ-medial axis (DMA), introduced in our recent work [9]. It proposes a fast algorithm that concurrently yields the construction of the discrete medial axis and the suppression of unwanted branches up to a single pruning parameter. Comparative results (Fig. 1, [9]) have shown the effectiveness of the DMA comparing to alternative approaches. Additionally, as will be discussed later in this paper, the value of the parameter δ can be deduced from the system properties, which will allow the whole topology extraction to be automatically processed.

(a) (b) (c)

Fig. 1: Medial axis pruning using: a) the Integer Medial Axis (IMA) [7], b) the γ-Integer Medial Axis (GIMA) [7], and c), the δ-Medial Axis (DMA) [9]. For images b) and c), the pruning parameter has been optimally set so that the resulting skeleton is homotopically equivalent to the shape (topology preserving).

Let $dt(\mathbf{x},\partial X)$ (also noted $dt(\mathbf{x})$ when the context is clear) be the distance transform of $\mathbf{x}$, corresponding to the minimal distance between $\mathbf{x}$ and ∂X: $dt(\mathbf{x}) = d(\mathbf{x}, ft(\mathbf{x}))$. Given $\delta \in \mathbb{R}^*_+$, we define [9] the δ-medial axis (DMA) of X as the set of points $\mathbf{x} \in X$ that verify the following conditions:

$$\exists \mathbf{y} \in X, d(\mathbf{x},\mathbf{y}) = 1 \ s.t. \begin{cases} \exists \mathbf{z} \in [ft(\mathbf{x}), ft(\mathbf{y})]; dt(\mathbf{z}) > \delta \\ \& \\ dt(\mathbf{x}) \geq dt(\mathbf{y}). \end{cases} \quad (4)$$

The first condition states that $\mathbf{x}$ and/or $\mathbf{y}$ belongs to the DMA if a disc of radius δ totally inscribed in X can be centered on a point $\mathbf{z}$ lying along $[ft(\mathbf{x}), ft(\mathbf{y})]$. While the GIMA [7] only considers the distance between two feature transforms, this first condition induces shape constraints in the neighborhood of the corresponding boundary points, leading to a better pruning. Similarly to the previous definition, the second condition selects the closest point to the real medial axis, and includes both if the points $\mathbf{x}$ and $\mathbf{y}$ are equally close.

III. Adapted skeleton for omnidirectional images

In term of shape analysis, the DMA produces a good approximation of the real medial axis in the image domain.

In our context however, it does not handle the anamorphosis of the projective catadioptric system where an image is a highly non linear distorted perspective of the real scene.

In the omnidirectional vision system, the imaging process is based on a two-step projection: Firstly, 3D rays are reflected by the mirror surface following the Snell's law at the intersection points; then, the reflected rays are acquired by the CCD camera into an image. The resulting projection is highly nonlinear and non uniform in the image domain. The used metrics for the medial axis computation should consequently be revisited. We propose here to define a conformal metric for feature and distance transforms that integrates those deformations and thus allows to construct an image skeleton in correlation with the 3D scene.

A. Projective model and omnidirectional image formation

Let $f : \mathbf{M} \in \mathbb{R}^3 \mapsto \mathbf{m} \in \mathbb{R}^2$ be a projection operator defined as $f([X, Y, Z]^t) = [x, y]^t$ mapping a 3D scene point to a point on the normalized image plane (see Fig. 2).

Using the equivalence sphere model [6], [1], that is a unifying projection model covering central catadioptric systems, the following equations hold:

$$f_s : \mathbf{M} \in \mathbb{R}^3 \mapsto \mathbf{M}_s \in \mathbb{S}^2, s.t. \begin{cases} X_s = X/\rho \\ Y_s = Y/\rho \\ Z_s = Z/\rho \end{cases} \quad (5)$$

$$f_{im} : \mathbf{M}_s \in \mathbb{S}^2 \mapsto \mathbf{m} \in \mathbb{R}^2, s.t. \begin{cases} x = X_s/(Z_s + \xi) \\ y = Y_s/(Z_s + \xi) \end{cases} \quad (6)$$

where $\rho = \sqrt{X^2 + Y^2 + Z^2}$, $\mathbf{M}_s$ is the projection of the 3D point $\mathbf{M}$ on the unitary sphere $\mathbb{S}$ and ξ is determined by the geometry of the mirror.

Combining equations (5) and (6), the projection operator f could be written:

$$f \equiv f_{im} \circ f_s : \mathbf{M} \in \mathbb{R}^3 \mapsto \mathbf{m} \in \mathbb{R}^2, s.t. \begin{cases} x = X/(Z + \rho\xi) \\ y = Y/(Z + \rho\xi) \end{cases} \quad (7)$$

B. Adapted metric for Distance Transform and Feature Transform

In the standard DMA formulation, both distance and feature transforms are computed based on the euclidian distance in the image domain. We propose to define an adapted metric on the omnidirectional image plane that preserves the euclidian distance of the 3D scene. This is a critical problem in the general case, and we will consider here the case of known altitude (Z is constant) for the practical use of computing the topology of the free space. This is a minor approximation since all the considered points belongs to X or ∂X, i.e. the ground (supposed to be plane) and its boundaries.

Given two points $\mathbf{M} = (X, Y, Z)^t$ and $\mathbf{N} = (X', Y', Z')^t$ in the free space and their corresponding projective points $\mathbf{m} = (x, y)^t$ and $\mathbf{n} = (x', y')^t$ in the normalized image plane, we define the adapted metric $\tilde{d}(\mathbf{m}, \mathbf{n})$ as:

$$\tilde{d}(\mathbf{m}, \mathbf{n}) = \alpha d_E(\mathbf{M}, \mathbf{N}) = \alpha d_E(f^{-1}(\mathbf{m}), f^{-1}(\mathbf{n})) \quad (8)$$

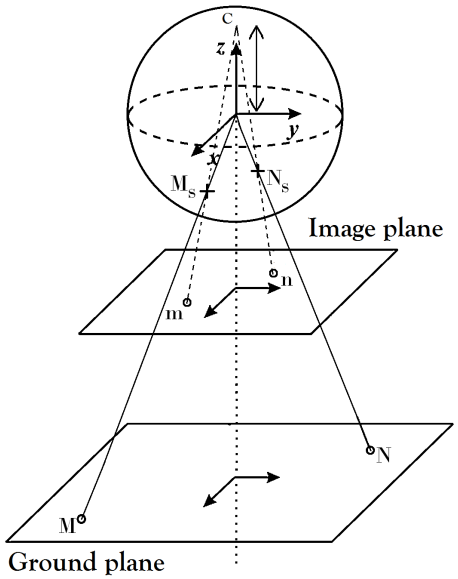


Fig. 2: Representation of the projection model that links two 3d ground points $\mathbf{M}$ and $\mathbf{N}$ to their projective image points $\mathbf{m}$ and $\mathbf{n}$, using the equivalence sphere model [1], [6]. Note that the metric scale is not observed to have a better understanding.

with α a scale factor. The inverse projection operator f^{-1} is expressed by:

$$f^{-1} : \mathbf{m} \mapsto \mathbf{M} = f_s^{-1} \circ f_{im}^{-1}(\mathbf{m}). \quad (9)$$

Noting that the operator $f_{im} : \mathbb{S}^2 \mapsto \mathbb{R}^2$ is a bijective transformation that maps the unit sphere into the image plane, its inverse can easily be computed:

$$\mathbf{M}_s = f_{im}^{-1}(\mathbf{m}) : \begin{cases} X_s = A(\mathbf{m})x \\ Y_s = A(\mathbf{m})y \\ Z_s = A(\mathbf{m}) - \xi \end{cases} \quad (10)$$

with $A(\mathbf{m}) = \dfrac{\xi + \sqrt{1 + (1 - \xi^2)(x^2 + y^2)}}{x^2 + y^2 + 1}$

Assuming the free space to be parallel to the normalized image plane, let consider $\mathbf{M}$ and $\mathbf{N}$ to belong to the free space with known (constant) Z ($Z = H$). The second inverse projection operator can thus be deduced from Eq. (5):

$$\mathbf{M} = f_s^{-1}(\mathbf{M}_s) : \begin{cases} X = HX_s/Z_s \\ Y = H\ Y_s/Z_s \\ Z = H \end{cases} \quad (11)$$

Finally, by combining Eq. (10) and (11), the inverse projection operator is written:

$$\mathbf{M} = f^{-1}(\mathbf{m}) : \begin{cases} X = HA(\mathbf{m})x/(A(\mathbf{m}) - \xi) \\ Y = HA(\mathbf{m})y/(A(\mathbf{m}) - \xi) \\ Z = H \end{cases} \quad (12)$$

The adapted metric in the omnidirectional image plane is then expressed, setting the scale factor $\alpha = 1/H^2$, by:

$$\tilde{d}(\mathbf{m},\mathbf{n})^2 = \frac{A^2(\mathbf{m})}{(A(\mathbf{m})-\xi)^2}(x^2+y^2) + \frac{A^2(\mathbf{n})}{(A(\mathbf{n})-\xi)^2})(x'^2+y'^2) \\ -2\frac{A(\mathbf{m})A(\mathbf{n})}{(A(\mathbf{m})-\xi)(A(\mathbf{n})-\xi)}(xx'+yy') \tag{13}$$

with:

$$A(\mathbf{m}(x,y)) = \frac{\xi + \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} \\ A(\mathbf{n}(x',y')) = \frac{\xi + \sqrt{1+(1-\xi^2)(x'^2+y'^2)}}{x'^2+y'^2+1} \tag{14}$$

The feature transform and distance transform are now expressed using the adapted metric formulation:

$$\widetilde{FT}(\mathbf{x}) = \left\{ \mathbf{y} \in \partial X | \forall z \in \partial X, \tilde{d}(\mathbf{x},\mathbf{y}) \leq \tilde{d}(\mathbf{x},z) \right\} \\ \widetilde{dt}(\mathbf{x}) = \tilde{d}(\mathbf{x}, \widetilde{ft}(\mathbf{x})) \text{ with } \widetilde{ft}(\mathbf{x}) \in \widetilde{FT}(\mathbf{x}) \tag{15}$$

The induced distance map in the omnidirectional image for the ground points, is depicted in Fig. 3.

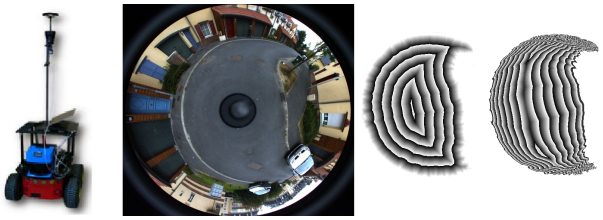


Fig. 3: Adapted metric for skeleton computation. *From the left to the right: i)* The omnidirectional vision system of the robot. *ii)* Input image captured by the system. *iii)* The Euclidian distance transform map dt of the ground points. *iv)* The distance transform map $\widetilde{dt}$ obtained using the adapted metric for omnidirectional images.

C. Computing Omnidirectional δ-Medial Axis

By extension, the omnidirectional δ-medial axis (ODMA) is formed by the set of points $\mathbf{x} \in X$ that verify:

$$\exists \mathbf{y}, d_E(\mathbf{x},\mathbf{y}) = 1 \; s.t. \begin{cases} \exists \; \mathbf{z} \in [\widetilde{ft}(\mathbf{x}), \widetilde{ft}(\mathbf{y})], s.t. \; \widetilde{dt}(\mathbf{z}) > \tilde{\delta} \\ \& \\ \widetilde{dt}(\mathbf{x}) \geq \widetilde{dt}(\mathbf{y}). \end{cases} \tag{16}$$

Note that the parameter $\tilde{\delta}$ being a distance threshold, it should also be expressed in the new metric. We choose $\tilde{\delta}$ to be correlated to the size of the robot. It is computed from the omnidirectional image by using the size of the blind spot (centre of the image) that encapsulates the sensor and robot body projection. This approximation is sufficient to automatically set $\tilde{\delta}$ to the radius (in the omni-adapted metric) of the circle encompassing the blind spot.

The DMA algorithm [9] produces the skeleton topology in linear time, but requires to be applied on a regular grid (since it is based on the algorithm of Hirata [8]). The adaptation to omnidirectional images is not trivial because the neighboring relation is nonlinearly modified by the induced metric. Clearly, the distance between two-grid points depends on their position in the image.

We instead use Danielsson's 8SED (8-neighborhood Squared Euclidian Distance) [5], a raster scanning algorithm that builds in linear time the approximated Euclidian distance map of a given shape (the generated error is largely discussed in [5]). Based on the assumption that each grid point shares its closest background point with at least one of its 8-neighbors (source of the approximation since this statement is not always true), the general idea is to propagate the distance to those closest points in two complete scans of the image. We adapt Danielsson's 8SED to compute the adapted metric for the image ground points and the corresponding feature $\widetilde{ft}$ and distance $\widetilde{dt}$ transforms. The general algorithm (using the adapted metric) is presented in Alg. 1.

By reconsidering the $\tilde{d}$ metric expression in Eq. (13), and noting $B(\mathbf{m}) = A(\mathbf{m})/(A(\mathbf{m})-\xi)$ and $r^2_{\mathbf{m}} = x^2+y^2$, we have:

$$\tilde{d}(\mathbf{m},\mathbf{n}) = B^2(\mathbf{m})r^2_{\mathbf{m}} + B^2(\mathbf{n})r^2_{\mathbf{n}} - 2B(\mathbf{m})B(\mathbf{n})(xx'+yy') \tag{17}$$

for each $\mathbf{m}(x,y)$ and $\mathbf{n}(x',y')$ in X. Note that $B(\mathbf{x})$ and $r^2_{\mathbf{x}}$ can be computed in an initialization step and stored for each point $\mathbf{x} \in X$ in a lookup table. That is very useful for real-time implementation.

Finally, the skeleton points are selected through a third pass of the image by testing the condition Eq. (16). Note that only few tests are made in practice: *i)* if $\tilde{d}(\widetilde{ft}(\mathbf{x}), \widetilde{ft}(\mathbf{y})) < 2\tilde{\delta}$, no scanning of the segment $[\widetilde{ft}(\mathbf{x}), \widetilde{ft}(\mathbf{y})]$ is performed; and *ii)* where appropriate, $\mathbf{z}$ is initiated to the midpoint of $[\widetilde{ft}(\mathbf{x}), \widetilde{ft}(\mathbf{y})]$, that in almost practical cases, maximizes the distance to the background. These optimizations allow the ODMA to perform in approximately linear time algorithm.

IV. Experiments and Comparative Results

A. Experimental setup

Our experimental platform is composed by a Pioneer 3-AT robot (see Fig. 3) equipped with a paracatadioptric omnidirectional sensor that captures color images with a resolution of 800×600, and positioned vertically at about 1.35m above the ground. The processing is carried online by a laptop mounted on the robot using Intel processor I7-2620m, able to compute all the processing (including image acquisition, free space segmentation, topology extraction and robot control) in less than 100ms.

The image segmentation (labeling each pixel as ground or background) is performed using a fast free space extraction algorithm presented in [10] and based on texture cues defined from local closest perception of the ground supporting the robot. The intuitive idea is to exploit them to propagate an active contour (easily initialized by encompassing the projection of the robot in the image), until it meets free space terminations. The method also integrates a false obstacle detector, and is by consequence well suited for road based scenarios (where white marks on the ground are omnipresent). The extracted free space in the image constitutes the starting input

Algorithm 1 $\tilde{dt}, \tilde{ft}$ maps computation

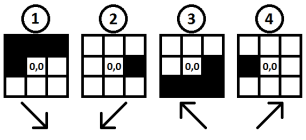
Input: Binary mask, $I = X \cup \overline{X}$ of size (row,column)
Output: Adapted distance and feature transform maps: $\tilde{dt}, \tilde{ft}$

for each $\mathbf{x} \in X$ $\quad \tilde{dt}(\mathbf{x}) = \infty, \tilde{ft}(\mathbf{x}) = NULL$ **end for**
for each $\mathbf{x} \in \overline{X}$ $\quad \tilde{dt}(\mathbf{x}) = 0, \tilde{ft}(\mathbf{x}) = \mathbf{x}$ **end for**

for i=0,row
  for j=0,column $\quad testPix((i,j), mask_1)$ **end for**
  for j=column,0 $\quad testPix((i,j), mask_2)$ **end for**
end for
for i=row,0
  for j=column,0 $\quad testPix((i,j), mask_3)$ **end for**
  for j=0,column $\quad testPix((i,j), mask_4)$ **end for**
end for

Procedure testPix(x, $mask$**)**
Input: Point $\mathbf{x}$, $mask$ neighbors to be considered

if $\mathbf{x} \in X$
  for each point $\mathbf{e} \in mask$
    $\mathbf{y} = \mathbf{x} + \mathbf{e}$
    if $\tilde{dt}(\mathbf{x}) > \tilde{d}(\mathbf{x}, ft(\mathbf{y}))$
      $\tilde{ft}(\mathbf{x}) = \tilde{ft}(\mathbf{y})$
      $\tilde{dt}(\mathbf{x}) = \tilde{d}(\mathbf{x}, ft(\mathbf{y}))$
    end if
  end for
end if



Danielsson's masks [5]

Algorithm 2 Skeleton points selection

Input: Adapted distance and feature transform maps $\tilde{dt}, \tilde{ft}$
Output: Omnidirectional δ-Medial Axis (ODMA)

for each $\mathbf{x}(i,j) \in X,\ \mathbf{y} \in \{(i-1,j),(i,j-1)\}$
  if $\tilde{d}(\tilde{ft}(\mathbf{x}), \tilde{ft}(\mathbf{y})) > 2\delta$
    if $segmentSearch(\tilde{ft}(\mathbf{x}), \tilde{ft}(\mathbf{y}))$
      if $\tilde{dt}(\mathbf{x}) \geq \tilde{dt}(\mathbf{y})$, add $\mathbf{x}$ to ODMA **endif**
      if $\tilde{dt}(\mathbf{x}) \leq \tilde{dt}(\mathbf{y})$, add $\mathbf{y}$ to ODMA **endif**
    end if
  end if
end for

Procedure segmentSearch($\tilde{ft}(\mathbf{x})$**,** $\tilde{ft}(\mathbf{y})$**)**
Input: Two extremities of the studied segment
Output: *true* if $\exists \mathbf{z} \in [\tilde{ft}(\mathbf{x}), \tilde{ft}(\mathbf{y})]\ s.t.\ \tilde{dt}(\mathbf{z}) > \delta$, *false* otherwise

for each $\mathbf{z} \in [\tilde{ft}(\mathbf{x}), \tilde{ft}(\mathbf{y})]$ (starting from the middle)
  if $\tilde{dt}(\mathbf{z}) > \delta$
    return true; // one point is sufficient: the procedure ends
  end if
end for
return false // no point verifies the condition.

for the proposed work. We refer the reader to [10] for more details on this algorithm.

Based on the extracted skeleton, a simple decision process is implemented to guide the robot following it. Robot control is referenced on the instantaneous perceived skeleton in the omni-image. In case of multiple skeleton branches, the robot selects the nearest branch to the left, since at this state of the work, no knowledge about the global topology is used.

Experiments were led in various indoor and outdoor real environments. We here focus on an outdoor scene where the robot, autonomously, traveled a total distance of more than 350m. We compare on the same sequence the topologies obtained using the standard δ-medial axis (where δ is manually set to maximize the performances of the algorithm, i.e. to limit noisy branches while extracting the topology information as far as possible from the robot), and the adapted version proposed above (as explained earlier, the value of the parameter is automatically set using image information, and thus, not optimized for the sequence).

Samples of the sequence and results where interesting topologies occur (corridor-like, crossroads from two different angles and curve trajectory) are presented in Fig. 4. As can be observed in the second and fourth rows, both versions produce a topology coherent with the extracted ground space in the image. At a first glance, one could even think that the standard method gives better results since undesired noisy branches appear at the end of the main branches for the adapted method (this is due to the lack of resolution for metric computation near the line horizon). However, for a better visualization, ground boundaries and extracted skeleton points are projected on the ground plane using intrinsic parameters (see third and last row of Fig. 4).

The immediate observation is that the standard method doesn't fit to the expected results, while the adapted version is much more closer to the exact scene Voronoi diagram. This is clearly a problem for safe navigation applications, since the robot will not follow the safest route (this is especially visible in the curve topology (Fig. 4). Moreover, a safe navigation is not the only purpose of the local topology. For instance, one may want to detect topological nodes (3D positions where several paths meet), which is impossible with the standard version, since the 3D projection of such detected points will move when the point of view changes. To highlight this phenomenon, figure 5 shows the superposition on the ground plane of several local topologies extracted at different positions of the same crossroad. While the adapted skeletons are almost superposed, one can easily observe that this is absolutely not the case for the standard version.

To prove the stability of the method in various situations, we encourage the reader to refer to the attached video also available at this link:

http://home.mis.u-picardie.fr/~marie/ICRA14/icra14.wmv

It presents the resulting skeletons computed by both non adapted and adapted versions for more that 100m of navigation and demonstrates the effectiveness of our approach and the relative stability of the extracted skeleton and its locus between consecutive images. It also shows that the quality of the ground segmentation is crucial. Minor deformations have almost no impact on the computed skeleton, but when huge deformations occur (missing parts of the ground or wrongly adding non ground regions), the resulting shape is no longer representative of the free space, and by consequence, the extracted topology might lead to dangerous or at least undesired navigation behaviors.

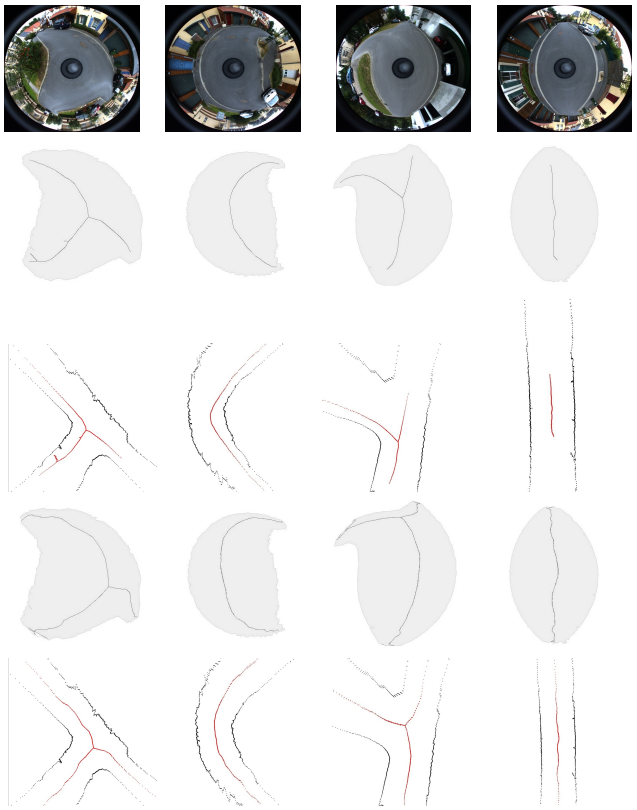


Fig. 4: Comparative results for standard and omnidirectional δ-MA. First row: Input images chosen for their topological interest. Second and third rows: Medial axis obtained with the standard metric and its projection on the ground plane. Fourth and fifth rows: Medial axis obtained with the adapted omnidirectional image metric and its projection on the ground plane.

V. Conclusion

In this paper, we have presented an original solution to model the topology of the environment by a robot using only the visual information provided by an omnidirectional camera. This work is based on the δ-medial axis (DMA), that is an efficient skeleton extraction algorithm for binary shapes. We have proposed an adapted version (ODMA) that integrates the distortions involved by the catadioptric projection in order to produce a skeleton in the image that fits to the Voronoi diagram of the 3D scene. The method uses a single parameter δ able to remove undesired branches, and is automatically set from the system characteristics provided by the image. Using efficient algorithm solutions and several optimizations, the skeleton (including the detection of the navigable space in the image), is computed fast enough (more than 10Hz) to be used online by a robot for autonomous tasks (safe navigation, map building, ...). Through experimental and comparative results, we highlight the use of the ODMA compared to a simple image based skeleton algorithm, regarding the quality of the extracted topology and its stability over time.

There are a number of areas to investigate in our future research. We intend to further work on the decision process to make complete autonomous exploration and on the fusion of successive skeletons for topological mapping purposes.

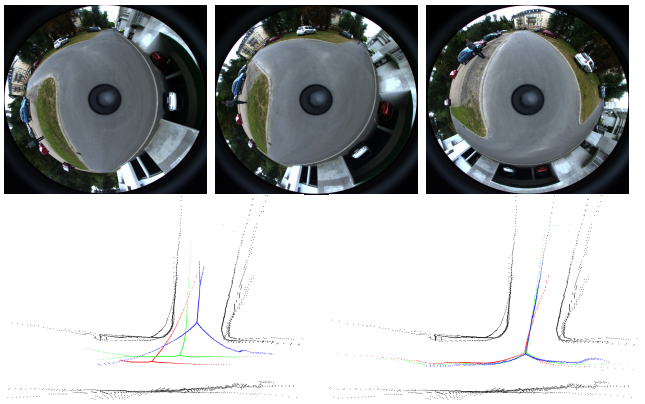


Fig. 5: Top: Three different views of the same crossroad. Bottom: Resulting skeletons for the standard version (left) and the omnidirectional adapted one (right).

References

[1] J. P. Barreto and H. Araujo, "Issues on the geometry of central catadioptric image formation," in *In CVPR*, 2001, pp. 422–427.

[2] M. R. Blas, M. Agrawal, A. Sundaresan, and K. Konolige, "Fast color/texture segmentation for outdoor robots," in *IEEE Int. Conf. Intelligent Robots and Systems*, 2008, pp. 4078–4085.

[3] H. Blum, "A transformation for extracting new descriptors of shape," in *Models for the Perception of Speech and Visual Form*, WWD, Ed. MIT Press, 1967, pp. 362–380.

[4] Z. Chen and S. Birchfield, "Qualitative vision-based path following," *Transactions on Robotics*, pp. 749–754, 2009.

[5] P. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol. 14, no. 3, pp. 227–248, 1980.

[6] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *Internation Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, 2001.

[7] W. Hesselink and J. Roerdink, "Euclidean skeletons of digital image and volume data in linear time by the integer medial axis transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2204–2217, 2008.

[8] T. Hirata, "A unified linear-time algorithm for computing distance maps," *Information Processing Letters*, vol. 58, no. 3, pp. 129–133, 1996.

[9] R. Marie, O. Labbani-Igbida, and M. Mouaddib, "The delta-medial axis: A robust and linear time algorithm for euclidian skeleton computation," in *to be published in ICIP*, 2013.

[10] P. Merveilleux, O. Labbani-Igbida, and M. Mouaddib, "Real-time free space detection and navigation using omnidirectional vision and parametric and geometric active contours," in *ICRA*, 2011, pp. 6312–6317.

[11] F. Paetzold and U. Franke, "Road recognition in urban environment," *Image and Vision Computing*, vol. 18, no. 5, pp. 377–387, 2000.

[12] C. Rasmussen, Y. Lu, and M. Kocamaz, "Appearance contrast for fast, robust trail-following," in *IEEE International Conference on Intelligent Robots and Systems*, 2009, pp. 3505–3512.

[13] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *Internation Journal of Computer Vision*, vol. 74, no. 3, pp. 237–260, 2007.

[14] S. Segvic, A. Remazeilles, A. Diosi, and F. Chaumette, "A mapping and localization framework for scalable appearance-based navigation," *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 172–187, 2009.

[15] A. C. Victorino, P. Rives, and J.-J. Borrelly, "Safe navigation for indoor mobile robots. part i: A sensor-based navigation framework," *International Journal of Robotics Research*, vol. 22, no. 12, 2003.