Model-Based Visual Self-Localization Using Gaussian Spheres

D. González-Aguirre, T. Asfour, E. Bayro-Corrochano and R. Dillmann

Abstract A novel approach for global self-localization is presented. It is based on a graph world-model and active stereo vision using density gaussian spheres. Evolved appearance-based object recognition components deliver noisy percept subgraphs which are filtered and fused into an ego-centered reference frame. In subsequent stages, the required vision-to-model associations are extracted by selecting ego-percept subsets in order to prune and match the corresponding world-model-subgraph. Ideally, these coupled subgraphs hold necessary information to obtain the model-to-world transformation, i.e. the pose of the robot. However, the estimation of the pose is not robust due to the uncertainties introduced when recovering euclidean metric from images and during the mapping from the camera to the ego-center. The approach models the uncertainty of the percepts with a radial normal distribution. This formulation allows a solution-optimization in a closed-form which not only derives the maximal density position depicting the optimal ego-center but also ensures the solution even in situations where pure geometric spheres might not intersect.

1 Motivation

Autonomous systems require the fundamental capability of self-localization in order to properly process, associate and interpret the incoming environmental sensor signals. A remarkable example of such systems are humanoid robots operating in structured environments also called *human-centered* environment, see Fig.1-a.

The use of a formal representation of the elements composing the surroundings and their interrelationships is needed to enable the robot to perform complex tasks

D. González-Aguirre, T. Asfour, R. Dillmann
Institute of Computer Science and Engineering, University of Karlsruhe, Haid-und-Neu-Strasse 7, Karlsruhe-Germany, e-mail: {gonzalez,asfour,dillmann}@ira.uka.de

E. Bayro-Corrochano
CINVESTAV, Av. Cientifica 1145, Zapopan-Jalisco, Mexico, e-mail: edb@gdl.cinvestav.mx

through the composition of multimodal skills accomplished through a perception-action cycle.

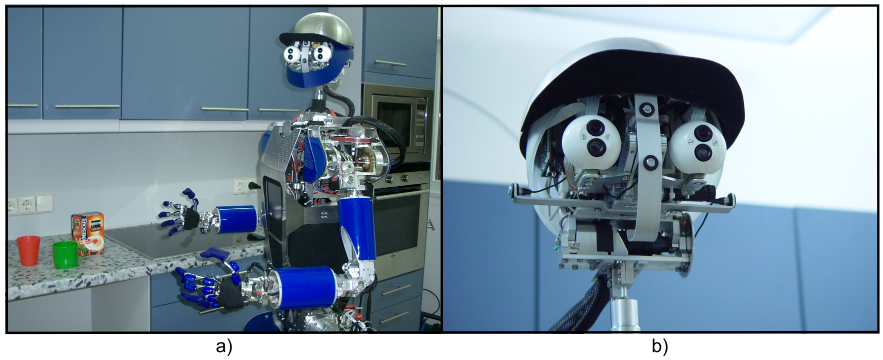

Fig. 1 a) The humanoid robot **ARMAR-IIIa** and its kitchen environment [1]. b) The active vision Karlsruhe humanoid head [2].

An effective mechanism to achieve the self-localization in these environments ought to profit from the intrinsic topological and geometric structure of the world by either constrainting the search within a tailored feature space or by extracting invariant properties of the world elements.

This mechanism has to sagaciously face many diminishing factors that complicate the self-localizing task, i.e. the granularity of the model, the nature of the sensors and the acquired uncertainty during the perception-recognition cycle.

This chapter presents a novel geometric and statistical approach to achieve the model based global self-localization using an active-vision sensing paradigm for humanoid robots, see Fig.1.

The global character of the localization concerns about the position and orientation (6D-*pose*) of the robot during the initialization which can also be stated as the kidnapping problem.

The natural and inherent usage of conformal geometric algebra [4] arise from the fundamental key idea of using conjuncted restriction subspaces in order to constraint and find the location of the robot. In this manner, the formulation profits from those sophisticated features of this powerful mathematical framework [5], for instance, the generalized intersection operator of geometric entities such as planes, lines, spheres, circles, pair of points and single points is the ideal instrument to attain the generation and validation of the ego-center location candidates of the robot.

This elegant treatment of subspaces helps to reduce the complexity of the percept-to-model matching by a computationally efficient, conceptually clear and consistent apparatus for expressing the intersection among the geometric primitives. In opposition to standard methods in linear algebra where usually a case-based procedure is applied to determine the intersection subspaces, the conformal geometric algebra provides a generalized mechanism, i.e. the meet operator [4][5].

2 Outline of Visual Self-Localization

The upper bar of the Fig.2 shows the three strata comprising the self-localization. First, the *physical space* encloses the real world were the robot is located. The *visual space* refers to the stratum of the process where the image information from the world is contained. Finally, the *world-model space* is a graph-based representation of the surroundings consisting of two sublayers; the geometric-level with the 3D vertices and their composition information and the topological-level describing the interrelation of object components.

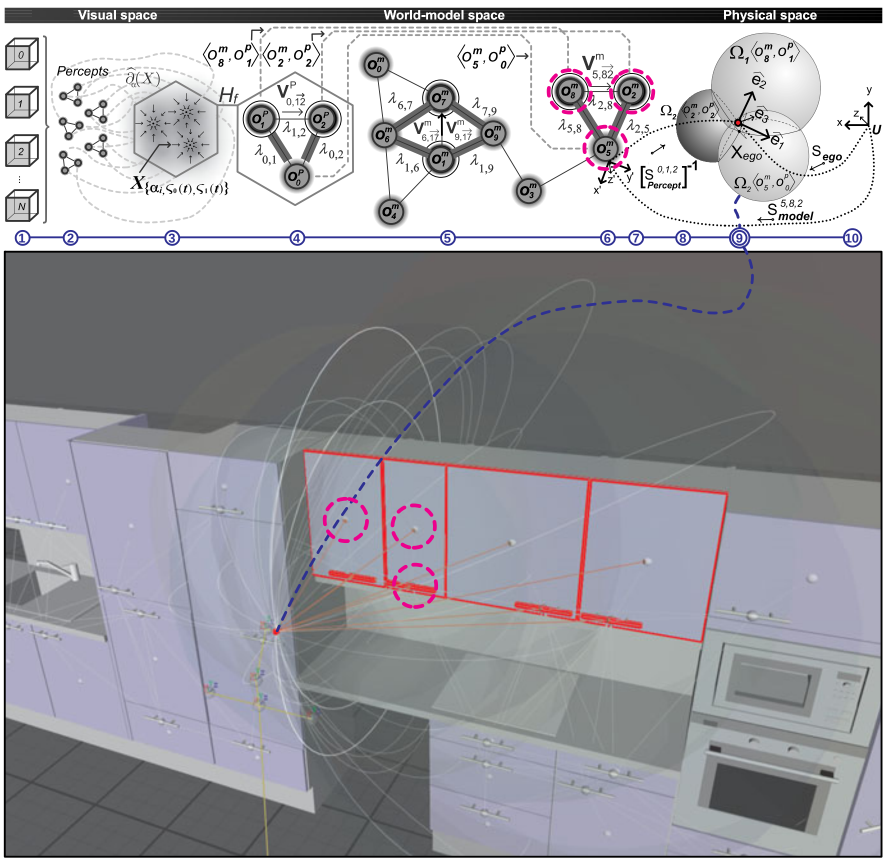

Fig. 2 Model-based visual self-localization approach, see [8]. 1) Appearance-based object recognition components. 2) Extracted percepts mapped into the ego-frame. 3) Multi-trial percepts fusion. 4) Fused ego-percepts with their corresponding world-model associations. 5) Proximity filtering for pruning purposes upon world-model. 6) Orientation filtering. 7) Hypotheses generation. 8) Hypotheses validation. 9) Geometric and statistical pose-estimation optimization. 10) Resulting pose.

Due to the nature of the problem, a model-based scheme, the global localization can be split into three sequential phases; visual acquisition of landmarks, data association for model matching and pose estimation-optimization.

2.1 Visual Acquisition of Landmarks

The active-vision perception and recognition components are responsible of delivering the position and orientation of the instances of those elements described in the world-model, see Fig.2.1-2. In opposition to previous approaches, the perception layer is not based on image saliencies or singularities such as *Harris* corners [6] or *SIFT* features [7] because these partially significant landmarks imply not only a burden during data association, but at certain point the humanoid robot utterly need to visually recognize the environmental elements in order to performing tasks.

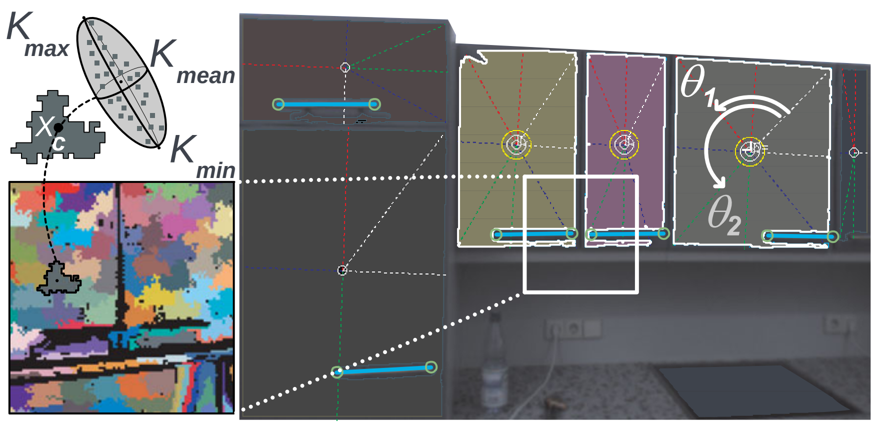

Fig. 3 Results of the class specific object recognition algorithms for door and door-handle, for a detailed description see [8].

In this way, the visually perceived and recognized instances (from now on *Percepts*) of those environmental objects provide not only useful information to perform actions but they also partially solve[1] the data association between the visual and model spaces. In this concrete context percepts are doors and door-handles in a building, see Fig.3. The advantage of using class based object recognition schema has been previously exploited [9]. In this way, faster and more robust methods can be applied.

In contrast, general feature approaches [10] lack of feature model association, besides offering poor reliability compared to those approaches mode for specific

[1] Up to the class instance association level.

domains. In this implementation, doors and door-handles were robustly recognized by means of gaussian classification over characteristic feature spaces extracted from class specific descriptors[2] of the eigenvectors[3] from color-segmented regions in stereo images, i.e. 2D Recognition. For an detailed description of the methods see [8]. Many specific recognition components may be added to improve the performance of the system at graph filtering by increasing the partition of the graph, i.e. reinforcing constraints and increasing pruning.

2.2 Data Association for Model Matching

There are two fundamental questions to be answered in order to properly solve the data association;

- How to fuse multiple percepts corresponding to the same world-element arising from multiple vantage points.
- How to match these fused-percepts against the world-model in order to compose the kinematic chain linking the selected perceptions to the world-model, i.e. the backwards transformation from the world to the robot, see Eq.5.

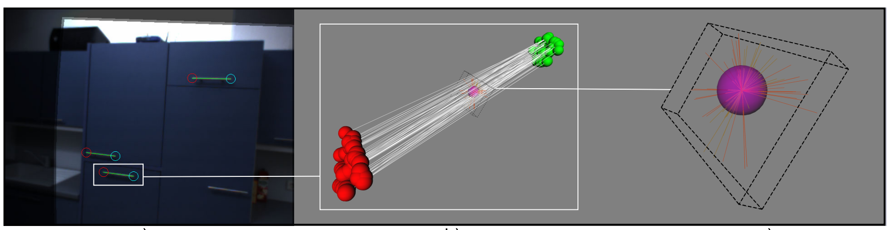

Fig. 4 a) Door-handle percepts recognized during scanning. b) Multiple percept corresponding to the same element in the world. c) Fused percepts into a stationary point $X_{\{\alpha_i,\varsigma_0(t),\varsigma_1(t)\}}$ of the underlying multimodal density function $\widehat{\partial_\alpha}(x)$, delineation set and its bounding box.

Percepts Fusion

Initially, a reference ego-space frame is defined, it is attached to a stable position of the humanoid robot, i.e. a kinematic frame of the robot which remains stationary during the visual scanning phase. Then, the time varying kinematic chain of transformations coupling the stereo vision system with the ego-frame is taken into account

[2] Specific tailored feature vector.

[3] From the covariance matrix of the clustered binary regions.

for the registration of the percepts. Subsequently, the percepts acquired during discrete steps of the scanning trajectory are mapped into the reference ego-frame, see Fig.4.

The underlying multimodal spatial density function,

$$\widehat{\partial_\alpha}(x) : \mathbb{R}^3 \mapsto \mathbb{R}$$

of the α-type percepts implies that stationary points

$$X_{\{\alpha_i, \varsigma_0(t), \varsigma_1(t)\}}.$$

These points are the locations of the α-modes, which describe the fused locations of the α-elements of the set. Percepts converging to $X_{\{\alpha_i, \varsigma_0(t), \varsigma_1(t)\}}$ constitute the fusion set, i.e. the cluster delineation in [12]. This is the key to properly fuse the multiple view percepts, see Fig.2.3.

These ideas are commonly used in the non-parametrical density estimation techniques as *Parzen* Windows [11] and Mean Shift [12]. The problem of estimating the bandwidth matrix and kernel type is coherently solved by using the geometric class-description of the percept, i.e. the inverse covariance matrix obtained from the 3D vertices of the geometric model.

The *Epanechnikov* kernel was chosen over the gaussian kernel because of its faster convergence. Although its faster convergence, the Epanechnikov kernel produces only negligible differences in the resulting delineation set compared with the results when using the gaussian kernel.

By exploiting these ideas, the multiple view perceptions are efficiently fused into a common reference space constituting the fused percepts set H_f, see Fig.2.4.

Fused-Percepts Matching

Previously merged landmarks are matched with the model by simultaneously trimming and coupling the elements of the world and those fused percepts, see Fig.2.5-6. In order to achieve this mechanism a graph based representation of the world is implemented whereas the fused percepts are arranged into a set of subgraphs according to their spatial distribution.

This coupling process requires to adequately incorporate the previous noisy fused-percept subgraphs as proper constraints to trim the model graph. In this way, the elements in the model which correspond to the selected acquired percepts remain active in the model space. Those elements which cannot satisfy the constraints are dismissed.

A selected percept subset could be partially matched against the model by using relative distances and orientations among them, i.e. removing elements which have no relative incidence within the perceived range of relative distances and orientations. This is the key idea of the *proximity* and *orientation* filtering.

For these purposes the world has been computationally modeled with two levels of abstraction. The first one describes the geometric composition of the elements and their relative pose. This is basically a CAD[4] structure. On this level the entities are data arrangements with information concerning 3D vertices and their composition describing geometric primitives. In the second level the latter structures compose instances of *object-model*[5] O_i^m with attributes, e.g. identifier, type, size and pose.

The collection of object-model instances constitutes the *node set* ν, whereas the *link set*

$$\Lambda \subset \{O_i^m \times O_j^m \ : \ O_i^m, O_j^m \in \nu, \ i > j, \ ||X_i - X_j|| < \zeta\}$$

depicts the connections $\lambda_{i,j}$ formed by all object model instances which relative distance[6] falls below ζ.

Proximity Filtering

When filtering links in the world-model graph, noise is taken into account in the form of deviation parameter ε_i function of the distance between the perceived-recognized objects[7] O_i^{pf}

$$\varepsilon_i = \frac{1}{\zeta}(||X_i^f - C_L||)^2 \tag{1}$$

with location X_i^f and center of the left camera C_L [14]. The result of the proximity filter is a set of links

$$\psi_{\{\alpha,\beta,\phi,\tau\}} \subset \Lambda$$

connecting nodes of type α to nodes type β, e.g. door to door-handle, which are separated by a distance ϕ with an error-tolerance

$$\tau = max_{k\in\Theta}(\varepsilon_k),$$

where Θ denotes the subset of recognized objects of both types:

$$\psi_{\{\alpha,\beta,\phi,\tau\}} \subset \{O^m_{(i,\alpha)} \times O^m_{(j,\beta)} \ : \ (\phi - ||X_i - X\,j||) < \tau\}.$$

The *active link set* consists of nodes from the intersection of those q proximity filtering partial results

---

[4] Coin3D: www.coin3d.org

[5] Note that the ”*m*” superscript emphasizes the fact that this is a model object instance.

[6] The magnitude of the threshold ζ corresponds to the maximal length of the 3D-FOV, see [13].

[7] Note that the ”*f*” superscript emphasizes the fact that this is a fused percept instance.

$$\psi_{act} := \bigcap_{i}^{q} \psi_{\{\alpha_i,\beta_i,\phi_i,\tau_i\}}.$$

Each filtering stage performs a strong reduction of the cardinality of the set ψ_{act} because those remaining nodes are tightly constrained, i.e. nodes should have neighbors with restricted types at constrained distance ranges. Fast performance was achieved by using a distance-lookup table when filtering only previously selected nodes.

Orientation Filtering

A more powerful, but computational expensive, technique to reduce the nodes within ψ_{act} is attained by accepting only those nodes which incidences have a relative pose. In this sense, the definition of the frame transformation has to be consistent while considering the noisy nature of the percept as follows:

First, three non-collinear elements are selected

$$O_i^{p_f}, O_j^{p_f} \text{ and } O_k^{p_f} \in H_f$$

then a frame is specified

$$S_{Percept}^{i,j,k} = [R_{Percept}^{i,j,k}, X_i^f]$$

relative to the ego-perception frame[8],

$$\widehat{\delta_1} = \frac{X_j^f - X_i^f}{||X_j^f - X_i^f||},\quad \widehat{\delta_2} = \frac{\left[\widehat{\delta_1} \wedge (X_k^f - X_i^f)\right]^*}{\left|\left|\left[\widehat{\delta_1} \wedge (X_k^f - X_i^f)\right]^*\right|\right|} \text{ and } \widehat{\delta_3} = \frac{\left[\widehat{\delta_1} \wedge \widehat{\delta_2}\right]^*}{||\left[\widehat{\delta_1} \wedge \widehat{\delta_2}\right]^*||},$$

which yields to

$$R_{Percept}^{i,j,k} = \left[\widehat{\delta^n} \cdot \widehat{e_n}\right]_{n=1\ldots3}.$$

Note these computations take place in $\mathbf{G}_{(3,0)}$, thus the dual of the wedge product of two vectors corresponds to the cross product in vector calculus.

Next, the relative displacement expressed on the frame of perception is computed

$$V_{i,\mathbf{jk}}^{p_f} = S_{Percept}^{i,j,k}(X_j^{p_f} - X_k^{p_f}),$$

such a vector merges the relative orientations of the three percepts in a signature-like consistent manner.

Therefore, it is possible to reject nodes which do not have a ”*similar*” displacement vector among two of the neighbors with corresponding type and proximity.

[8] Which orthonormal basis vectors are $\{\widehat{e_1}, \widehat{e_2}, \widehat{e_3}\}$

This *noisy-similarity* is quantified by the length and angle discrepancies μ and κ between the perception signature $V^p_{i,\mathbf{jk}}$ and the model signature $V^m_{u,\mathbf{uw}}$ vectors, expressed on the world-model $S^{u,w,v}_{model}$.

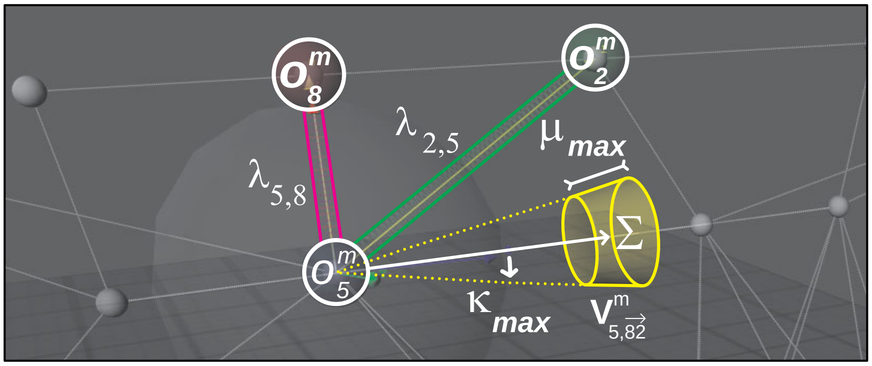

Fig. 5 World-model graph at pruning by means of proximity and orientation filtering. Example of accepted node O^m_5 with vector $V^m_{5,\mathbf{8},\mathbf{2}}$ inside Σ. Notice, the subspace Σ corresponds to the boolean subtraction of two spherical cones [3]. The aperture of the implicit cone depicts the noise parametrical tolerance of the orientation filtering, see Eq.3. The radii of both implicit spheres differ by μ_{max}, i.e the proximity filtering noise parametrical tolerance, see Eq.2.

Fig.5 shows the subspace Σ bounded by

$$||V^p_{i,\mathbf{jk}} - V^m_{u,\mathbf{vw}}|| < \mu_{max}, \tag{2}$$

$$\arccos(\widehat{V^p_{i,\mathbf{jk}}} \cdot \widehat{V^m_{u,\mathbf{vw}}}) < \kappa_{max}. \tag{3}$$

When filtering a node, the combinational burst is avoided by computing only subgraphs which link lengths falls into the range

$$(|V^p_{i,\mathbf{jk}}| - \mu_{max}) < ||[O^{p^f}_j, O^{p^f}_k]|| < (|V^p_{i,\mathbf{jk}}| + \mu_{max}).$$

2.3 Pose-Estimation Optimization

Previously extracted model subgraphs which simultaneously match the typed incidences and relative pose of those acquired *percepts subgraphs* embody the association coupling the *visual space*, *world-model* and *physical world*.

They simultaneously impose restraints which are the *geometric-compelling keys* to deduct the pose of the robot. Each association

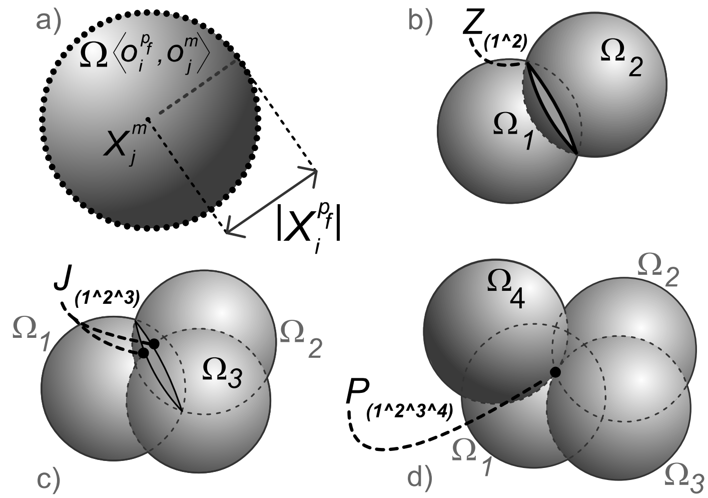

Fig. 6 a) Constrained-subspace embodies the surface on the sphere. b) Cooccurring constrained-subspaces depicting a circle. c) Three acting in conjunction constrained-subspaces yielding to a pair of points. d) Four constrained-subspaces yielding to a simultaneity-point.

$$\left\langle O_i^{p_f}, O_j^m \right\rangle$$

constraints the position of the robot to the subspace of all points which are $||X_i^{p_f}||$ units away from X_j^m.

This subspace is actually the surface on a sphere, i.e.

$$\underbrace{\Omega\left\langle O_i^{p_f}, O_j^m \right\rangle}_{Restriction\ Subspace} := X_j^m + \frac{1}{2}\underbrace{\left(||X_j^m|| - ||X_i^{p_f}||\right)}_{Perception-Model\ Matching} e_\infty + e_0 \in PK^3 \qquad (4)$$

centered at X_j^m with radius $||X_i^{p_f}||$, see Fig.6-a.

Note that the sphere in Eq.4 is a element of the conformal geometric space PK^3 which has the Clifford algebra signature $\mathbf{G}_{(4,1)}$ [4].

For a single percept this idea provides no benefit, but on second thought, when observing the same concept with two different percepts it turns out to be a very profitable formulation because the ego-center should yield in both constrained sub-spaces, meaning it has to be on the surface of both spheres at the same time.

Consider two *restriction spheres* simultaneously constraining the position of the robot,

$$\Omega_1\left\langle O_i^{p_f}, O_j^m \right\rangle \ \ and \ \ \Omega_2\left\langle O_k^{p_f}, O_l^m \right\rangle$$

they implicate that the position of the robot belongs to both subspaces. Thus, the restricted subspace is a circle, i.e. an intersection of spheres, see Fig.6-b,

$$Z_{(1\wedge 2)} = \Omega_1 \left\langle O_i^{p_f}, O_j^m \right\rangle \wedge \Omega_2 \left\langle O_k^{p_f}, O_l^m \right\rangle .$$

Following the same pattern, a third sphere Ω_3 enforces the restriction to a pair of points

$$J_{(1\wedge 2\wedge 3)} = Z_{(1\wedge 2)} \wedge \Omega_3 \left\langle O_r^{p_f}, O_s^m \right\rangle$$

i.e. circle-sphere intersection, see Fig.6-c. Finally, a fourth sphere Ω_4 determines the position of the robot, i.e. the intersection point from the latter pair of points, see Fig.6-d,

$$P_{(1\wedge 2\wedge 3\wedge 4)} = J_{(1\wedge 2\wedge 3)} \wedge \Omega_4 \left\langle O_t^{p_f}, O_h^m \right\rangle .$$

Latter concepts outline a technique which uses the previously partially matched elements of the world-model and process them by a geometric apparatus for generating the ego-center candidates. This apparatus uses the centers of the spheres within the model space and the radii from the fused-percepts, see Fig.2.6-8 and Fig.7.

The computational complexity of this *location hypotheses* management process is upper bounded by $\mathbf{O}(n^4)$, where n is the cardinality of the subset of percepts-spheres.

The amount of spheres n is by no means bigger than 6 while generating candidates, besides in rare cases the internal partial result of the intersection stages are densely populated. This could be easily seen when intersecting two spheres. The resulting circle occupies a smaller subspace which in successive stages meets only fewer remaining spheres. One important factor why there are less operations in this combinational computation is because the child primitives that result from the intersection of parent spheres should not be combined with their relatives avoiding useless computation effort and memory usage.

Hypotheses Generation

Each percepts subgraphs is used to produce the *zero-level set*, composed of spheres, see Fig.2.7,

$$\Phi_0 = \left\{ \Omega_\zeta \left\langle O_i^m, O_j^p \right\rangle \right\}_{\zeta = 1 \ldots n} .$$

These spheres are then intersected by means of the *wedge* operator $\wedge$ in an *upper triangular* fashion producing the *first-level set* Φ_1 containing circles.

The *second-level set* Φ_2 is computed by intersecting those circles with spheres from Φ_0 excluding those directly above. Then the latter resulting pair-points are intersected in the same way creating the highest possible stratum (*third-level set*) Φ_3, here the points resulting of the intersection of 4 spheres are contained.

Finally, elements of Φ_2 which have no descendants in Φ_3 and all elements on Φ_3 represent location hypotheses

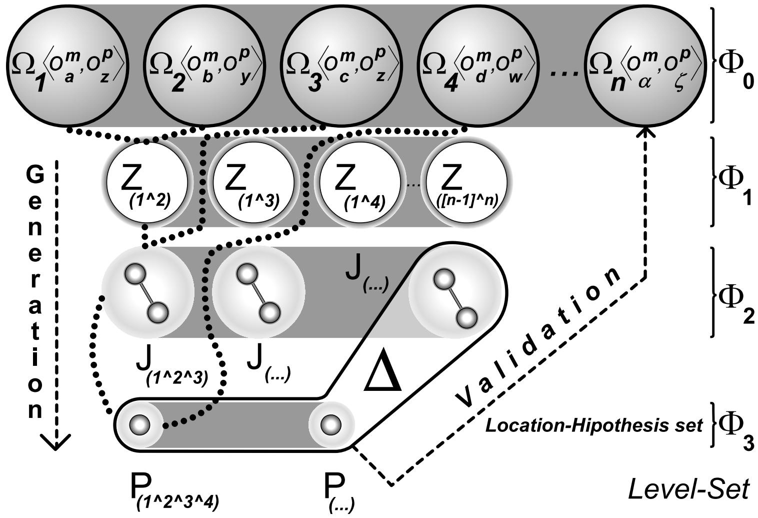

Fig. 7 Location hypotheses generation-validation mechanism systematically manages the location-hypotheses.

$$\Delta := \bigwedge_{\xi} \Omega_{\xi} \left\langle O_i^m, O_j^p \right\rangle.$$

Hypotheses Validation

Hypotheses are checked by selecting associations, see Fig.2.8,

$$\left\langle O_i^{p_f}, O_j^m \right\rangle$$

which were not considered when the current validating hypothesis was generated. In case there is more than one prevailing hypothesis, which rarely happens in non-symmetric repetitive environments, an active validation needs to take place selecting objects from the model and then localizing them in the visual space. The criterion to select the discriminator percept $D_{i,j}^m$ (priming instance) is the maximal pose difference between hypotheses pairs.

Ideal Pose Estimation

Once the location hypothesis has revealed the position of the robot X_{ego} (see Fig.2.10) the orientation S_{ego} is expressed as

$$\underbrace{S_{ego}}_{Self-Localization} = \underbrace{S_{model}^{u,w,v}}_{Model-Matching} \underbrace{[S_{Percept}^{i,j,k}]^{-1}}_{Visual-Perception} , \tag{5}$$

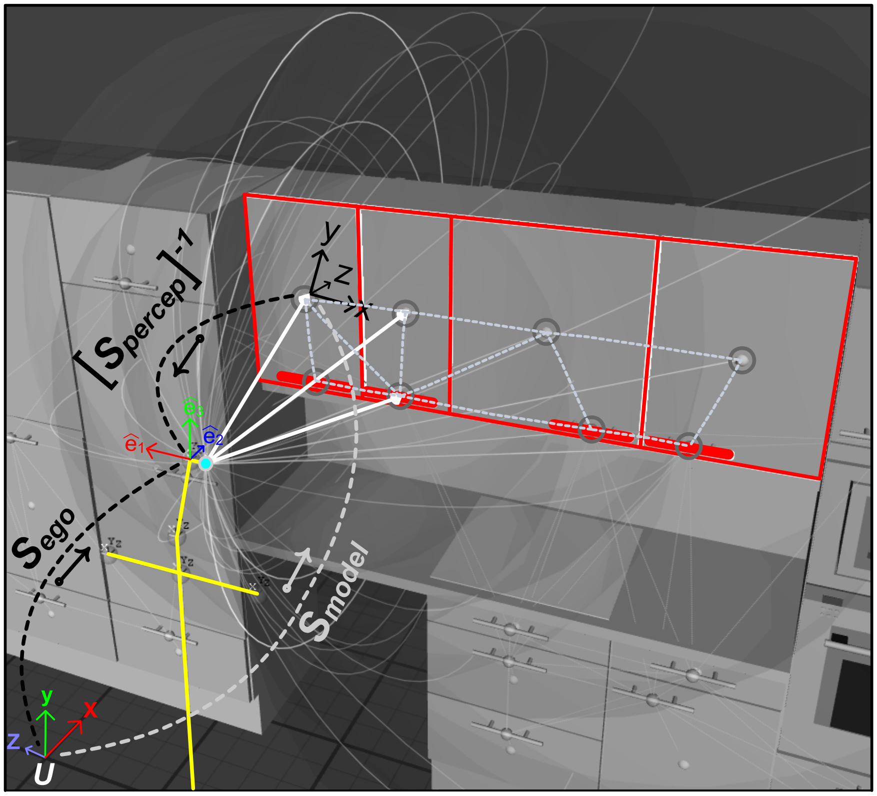

Fig. 8 Kinematic frames involved in the ideal visual self-localization. Notice the directions of the coupling transformations in order to reveal the frame S_{ego}.

which is actually the transformation from the kinematic chain that couples the world-model frame S_{model} (*forwards*) and the perception frame $[S_{Percept}^{i,j,k}]^{-1}$ (*backwards*), see Fig. 8.

There are situations where a variety of diminishing effects alter the depth calculations of the percepts in a way that the ideal pose calculation may not be robust or could not be assessed. The subsequent sections describe the sources and nature of the uncertainties, which are modeled and optimized by the proposed technique to find the location of the robot, i.e. the maximal probabilistic position.

3 Uncertainty

The critical role of the uncertainty cannot only strongly diminish the precision of the estimated pose, but it can also menace the existence of it by drawing away the intersection of the restriction subspaces, i.e. the spheres might not intersect due to numerical instability and errors introduced by the perception layer.

In order to sagaciously manage these conditions and other derived side effects, it is crucial to reflect upon the nature of the acquired uncertainties regarding this localization approach. There are two remarkable categorical sources of uncertainty, *image-to-space* and *space-to-ego* uncertainties.

3.1 Image to Space Uncertainty

Image-to-space uncertainty is obtained from the appearance-based vision recognition process. It starts with the pixel precision limitations, e.g. noise, discretization, quantization, etc. and ends with the error-limitations of the camera model and its calibration, e.g. radial-tangential distortion and intrinsic parameters [15]. This uncertainty could be modeled, according the central limit theorem [16] as a normal distribution where the variance is strongly related to the perceptions depth

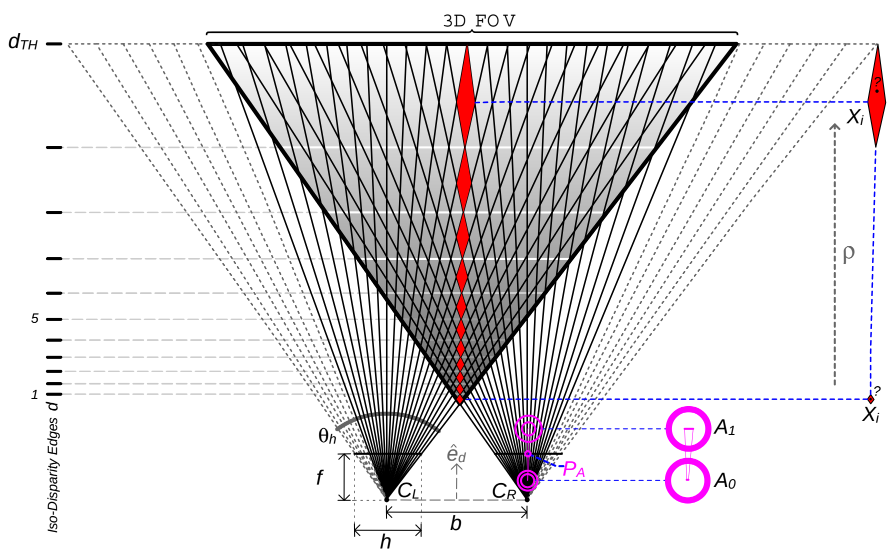

Fig. 9 The image-space uncertainty factors in a front-parallel configuration.

$$\rho_i = (x_i - C_L) \cdot \hat{e}_d, \tag{6}$$

i.e. the distance between camera center C_L and point in space x_i along the stereo rig normal vector $\hat{e}_d$, see Fig.9,

$$\sigma_i \cong \frac{1}{\zeta}\rho_i^2, \tag{7}$$

where $\zeta > 1 \in \mathbb{R}$ is an empirical scalar factor depending on the resolution of the images and the vergence angle of the stereo rig. This variance model arises from the following superposed facts: first, considering only the monocular influence in each camera of the stereo rig.

The surface patch A_i on the plane perpendicular to the optical axis of the camera imaged into a single pixel P_A grows linearly as function of the distance ρ_i

$$A_i = 4\rho_i \tan\left(\frac{\theta_h}{2h}\right)\tan\left(\frac{\theta_v}{2v}\right),$$

where θ_h and θ_v represent the horizontal and vertical angular apertures of the field of view, whereas h and v depict the width and height resolutions of the image, see Fig.9.

Consequently, the stereo triangulation has an additional effect during the estimation of the 3D position $\mathbf{M}_{stereo}(X_i)$ of a matched pair of points. The distance ρ_i affects the magnitude of the disparity d_i. Therefore, the precision of the pixel computations plays a decisive role, i.e. the 3D space points which are closer to the base line have wider disparities along the epipolar lines, meanwhile those points located after distance $\rho_{Th} > fb$ have a very narrow disparity, falling in subpixel domain $d < 1$, which results in inaccurate depth calculations.

This situation also produces a sparse distribution of the iso-disparity surfaces[17], meaning that the subspace contained between this surface-strata grows as

$$d_i = \frac{fb}{\rho_i}, \tag{8}$$

where the focal distance f and the base line size play relevant role in the measurement precision

$$b = ||C_L - C_R||.$$

Fig.9 shows the ideal front parallel case iso-disparity edges delineating the subspaces contained between two discrete steps in the disparity relation of the Eq.8.

In this manner, points contained within one of these subspaces produce the same discrete disparity when matching corresponding pixels. Hence, the location uncertainty ought to be proportional to the distance contained between iso-disparity surfaces. These two applied factors produce an uncertainty growing in an attenuated quadratic fashion, which is reflected in the model as a variance spreading in the same pattern reflexted upon Eq.6.

3.2 Space to Ego Uncertainty

The space-to-ego uncertainty is acquired while relating the pose of the percepts from the left camera frame to the ego-frame (head-base frame of the humanoid robot, see Fig.10-a).

It is caused by the physical and measurement inaccuracies, which are substantially magnified by projective effects, i.e. those almost negligible errors in the encoders and mechanical joints of the active head of the humanoid robot are amplified proportional to the distance ρ_i between the ego center and the location of the percept.

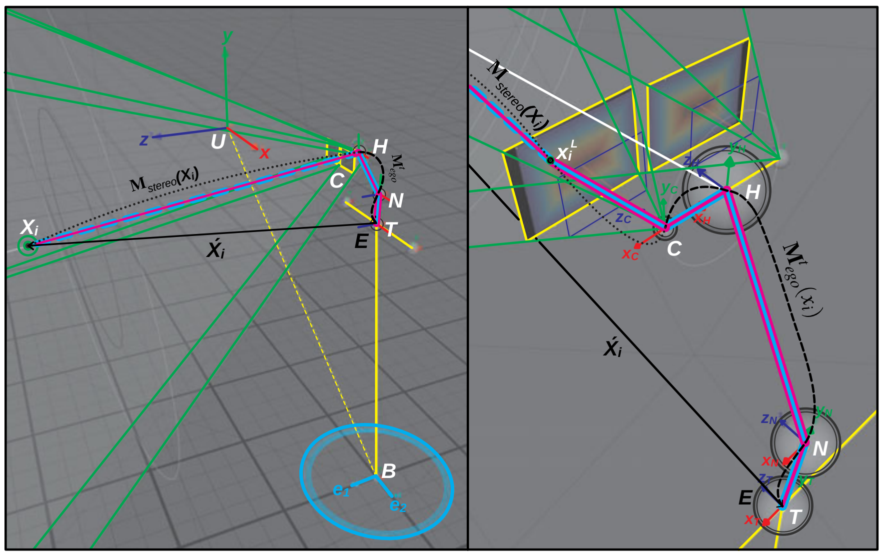

Fig. 10 The space-ego uncertainty acquisition process produced by the mapping of percepts from camera coordinates to the ego-frame. a) The whole transformation $\acute{X}_i = \mathbf{M}^t_{ego}(\mathbf{M}_{stereo}(X_i))$. b) The transformation $\mathbf{M}^t_{ego} = [T_{(t)}N_{(t)}HC_L]^{-1}$.

Fig.10-b shows the kinematic chain starting at x_i^L, the left camera coordinates of the space point X_i. Subsequently, the transformation from the left camera frame C_L to the shoulders base $T(t)$ passing through the eyes base H and neck frame $N(t)$, is given by

$$\acute{X}_i = \mathbf{M}^t_{ego}(x_i), \tag{9}$$

$$\mathbf{M}^t_{ego} = [T_{(t)}N_{(t)}HC_L]^{-1}, \tag{10}$$

where $\mathbf{M}^t_{ego}$ is the ego-mapping at time t. Here, the transformations $T_{(t)}$ and $N_{(t)}$ are time-dependant because they are active during the execution of the scanning strategy, see Fig.10-b.

4 Geometry and Uncertainty Model

Once the visual recognition components provided all classified percepts within a discrete step of the scanning trajectory, these percepts are mapped into the reference ego-frame using Eq.9. This ego-frame is fixed during the scanning phase. In this fashion all percepts from different trials are located in a static common frame, see Fig.10-b.

The unification-blending process done by the fusion phase simultaneously allows the rejection of those percepts which are far from being properly clustered and creates the delineation set which is later melted into a fused percept.

Next, the geometric and statistical phase which determines the position of the robot based on intersection of spheres is properly formulated by introducing the following gaussian sphere and its apparatus for intersection-optimization.

4.1 Gaussian Spheres

The considered restriction spheres Ω_i are endowed with a soft density function

$$\widehat{f}(\Omega_i, x) \quad \Omega_i \in PK^3,\ x \in \mathbb{R}^3 \mapsto (0,1] \in \mathbb{R}.$$

The density value decreases exponentially as a function of the distance from an arbitrary point x to the surface of the sphere Ω_i

$$S(x, X_i, r_i) = |(||x - X_i|| - r_i)|\,, \tag{11}$$

$$\widehat{f}(\Omega_i, x) = e^{\frac{-S(x,X_i,r_i)^2}{2\sigma_i^2}}. \tag{12}$$

Latter function depicts the non-normalized [9] radial normal distribution

$$\check{N}(\mu := \{x \mid \ker(S(x, X_i, r_i))\}, \sigma_i^2)$$

for x to be in the surface of Ω_i, i.e. the null space of $S(x, X_i, r_i)$. Note that here the variance σ_i refers to Eq.7.

The density of a point x in relation with a sphere Ω_i represents the non-normalized probability for the point x to belong to the surface of the sphere Ω_i. Obviously the maximal density is on the surface of the sphere itself.

[9] By the factor $\frac{1}{\sigma\sqrt{2\pi}}$.

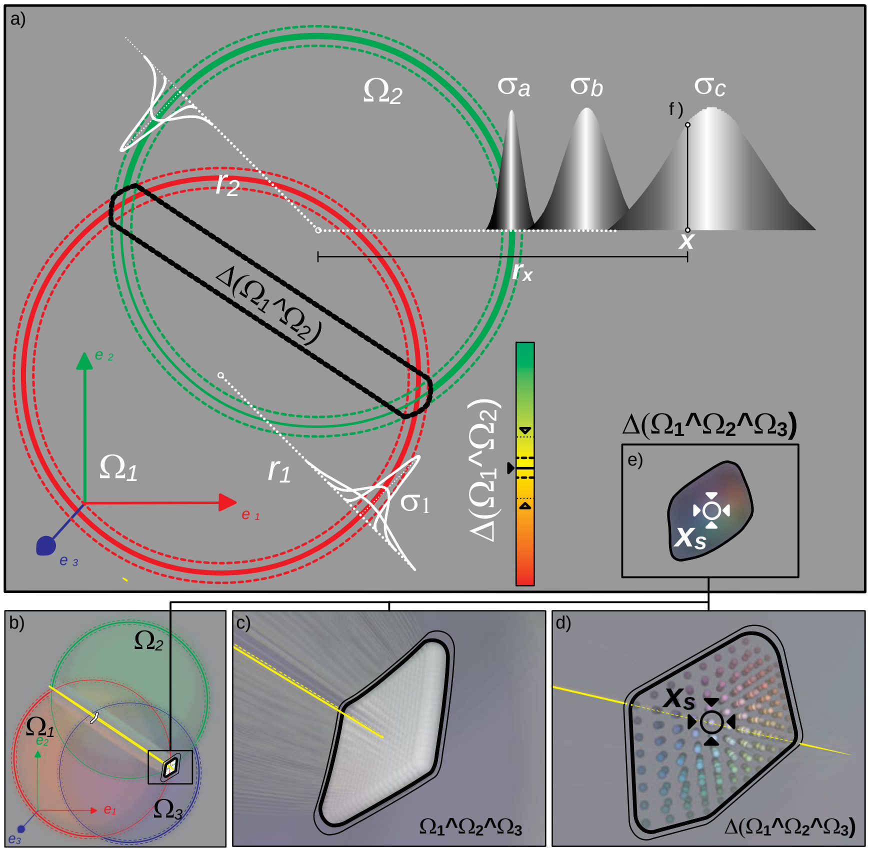

Fig. 11 Gaussian spheres meeting. a) Two gaussian spheres meeting $\Omega_1 \wedge \Omega_2$ describing a density-subspace $\Delta(\Omega_1 \wedge \Omega_2)$. b) Three gaussian spheres $\Omega_{i=1,2,3}$ meeting in two regions depicting a subspace $\Omega_1 \wedge \Omega_2 \wedge \Omega_3$. c) Detailed view of one of the previous subspaces. d) Discrete approximation of the maximal density location x_s. e) Details of the implicit density-space $\Delta(\Omega_1 \wedge \Omega_2 \wedge \Omega_3)$. f) Implicit radius r_x when estimating the density at position x.

It is necessary to propose an effective mechanism which applies intersections of restriction spherical subspaces as essential idea for finding the robot position. The nature of the applied intersection has to consider the endowed spatial density of the involved gaussian spheres.

In the following sections, the restriction spheres and their conjuncted composition properly model both uncertainties, allowing the meeting of spheres by finding the subspace where the maximal density is located, see Fig.11.

This could be better understood as an isotropic dilatation or contraction of each sphere in order to meet at maximal density of the total density function, see Fig.12 and Fig.13,

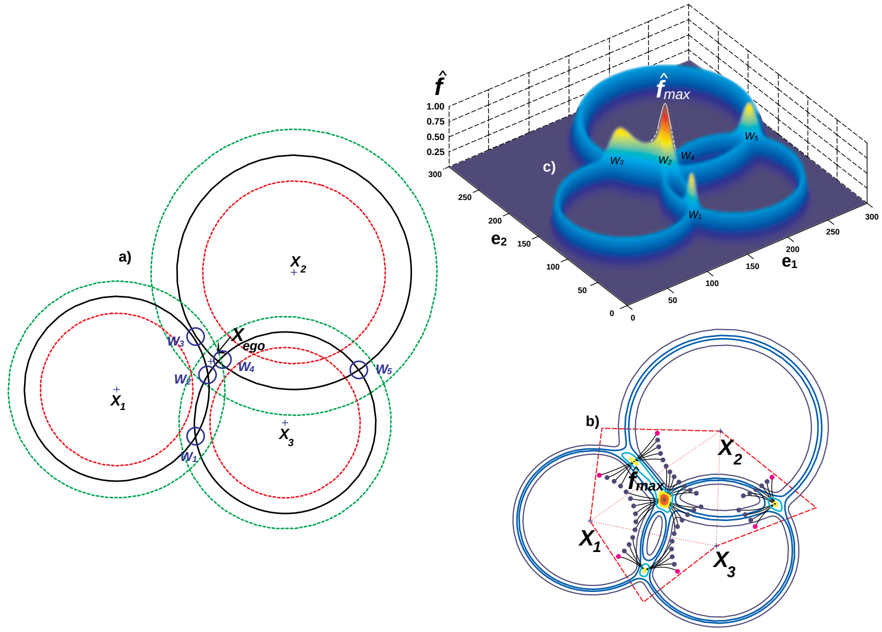

Fig. 12 Gaussian circles, i.e. 2D gaussian spheres. a) Three gaussian circles setup. b) The total accumulative density $\widehat{f_c}(x) = \sum_i^n \widehat{f}(\Omega_i, x)$ allows a better visualization of the composition of its product counterpart $\widehat{f_t}(x)$, see also Fig.13. c) Density contours with seeds and their convergence by means of gradient ascendant methods.

$$\widehat{f_t}(x) \longrightarrow (0,1] \in \mathbb{R},\ x \in \mathbb{R}^3 \tag{13}$$

$$\widehat{f_t}(x) = \prod_i^n \widehat{f}(\Omega_i, x). \tag{14}$$

Due to the geometric structure composed by n spheres, it is possible to foresee the amount of peaks and the regions $\mathbf{W}_s$ where the density peaks are located. Therefore, it is feasible to use state-of-the-art gradient ascendant methods [18] to converge to the modes using multiple seeds. These should be strategically located based on the spheres centers and intersection zones, see Fig.12.

Finally, the seed with maximal density represents the solution position x_s

$$x_s = argmax\ \widehat{f_t}(x). \tag{15}$$

However, there are many issues of this shortcoming solution. The iterative solution has a precision limited by the parameter used to stop the shifting of the seeds. In addition, the location and spreading of the seeds could have a tendency to produce undesired oscillation phenomena, under or oversampling and all other disadvantages that iterative methods present.

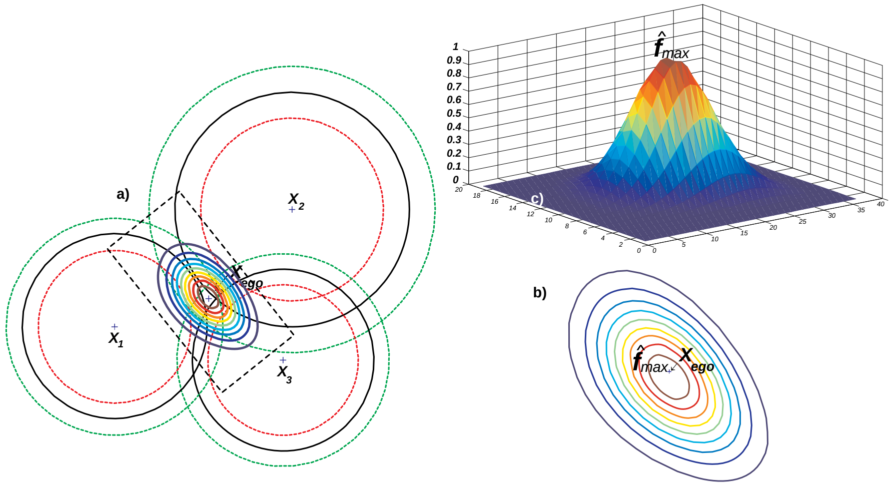

Fig. 13 The gaussian circles, i.e. 2D gaussian spheres. a) Three gaussian circles setup. b) The total density $\widehat{f}_t(x) = \prod_i^n \widehat{f}(\Omega_i, x)$. c) Density contours and ego-center X_{ego}, notice that the resulting distribution is not gaussian.

The optimization expressed by Eq.15 could be properly solved in a convenient closed-form. In order to address the solution x_s, it is necessary to observe the configuration within a more propitious space, which simultaneously allows an advantageous representation of the geometrical constraint and empowers an efficient management of the density, i.e. incorporating the measurements according their uncertainty and relevancy while avoiding density decay.

4.2 *Radial Space*

The keys to attain a suitable representation of the latter optimization resides in the exponent of the Eq.12. There, the directed distance from a point x to the closest one on the surface of the sphere is expressed by the Eq.11. When considering the total density function (see Eq.14) it unfolds the complexity by expressing the total density as a tensor product.

The inherent nature of the problem lies in the radial domain, i.e. the expression $S(x, X_i, r_i)^2$ is actually the square magnitude of the difference between the radius r_i and the implicit defined radius r_x between the center of the spheres X_i and the point in question x, see Fig.11-f. Hence, the optimization configuration can be better expressed in radial terms and the geometrical constraints restricting the relative positions of the spheres is properly and naturally uncluttered in the following sections.

4.3 Restriction Lines

Consider the case of two spheres Ω_1 and Ω_2, see Fig.14-a. Here, the radii of both spheres and the distance between their centers

$$\delta_{\overline{1,2}} = ||X_1 - X_2|| = \sqrt{-2(\Omega_1 \cdot \Omega_2)}$$

allow the formulation of the geometric restrictions, which ensure the intersection of the spheres in at least a single point P_χ.

These restrictions are expressed by the inequation line L_χ, which describes the radial configuration subspace represented by pairs of the form

$$P_\chi = [r_1, r_2]^T \in \mathbf{S}^2,$$

the intersection of spheres $\Omega_1 \wedge \Omega_2$, i.e. a circle with null radius, where the $\mathbf{S}^2$ refers to the radial configuration space of two spheres.

Notice in Fig.14-d the inequality line divides the configuration space into two regions, the half space partially holding the restriction imposed by the inequation line L_χ, however in this space there are still configurations which produce no intersection of spheres, in fact any configuration holding

$$r_2 \geq \delta_{\overline{1,2}} + r_1.$$

In order to prevent these degenerated configurations two additional restriction inequation lines arise, unveiled by following similar pattern.

In the same fashion, Fig.14-b shows the case where the minimal contact point P_β occurs, subject to

$$r_1 \geq \delta_{\overline{1,2}} + r_2.$$

In this configuration subspace, the sphere Ω_1 fully contains sphere Ω_2 and their surfaces intersect solely at P_β. Once again, in order to ensure at least this contact point, the fluctuation of the radii of both spheres is restricted by a linear relation expressed by the inequality line L_β.

The latter restriction actually happens in a symmetric manner by interchanging the roles from Ω_1 with Ω_2, resulting in a third restriction, i.e. the inequality line L_α, see Fig.14-c,d.

As a result, the configuration space is divided in four regions K_α, K_β, K_χ and K_{meet} all open except K_χ.

Only those configurations within the subspace K_{meet} represent non-empty intersections of the spheres, e.g. the point x_{conf} in Fig.14-d with

$$x_{conf} = [r_1', r_2']^T \in K_{meet}.$$

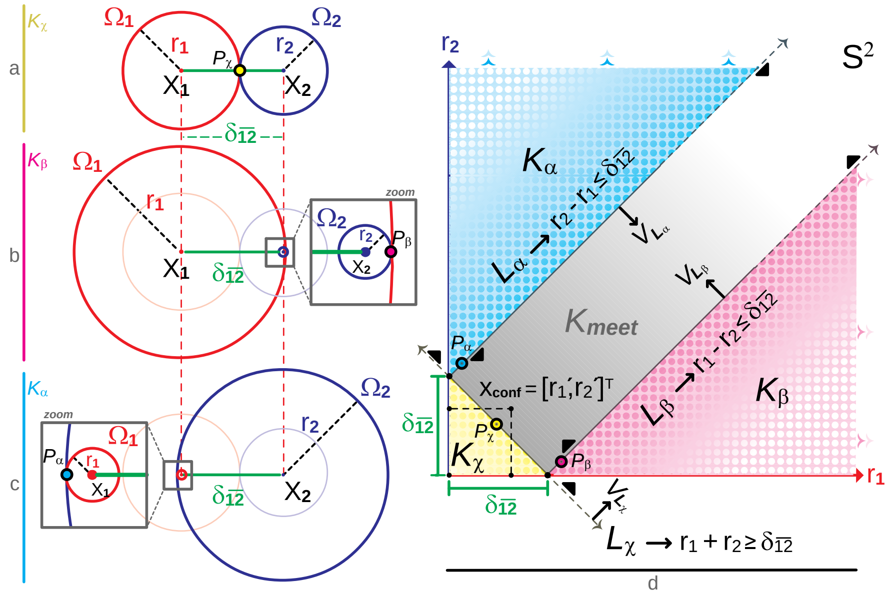

Fig. 14 The spheres intersection restriction lines derivation in the radial space $\mathbf{S}^2$. a) The line L_χ is the first restriction for ensuring non-empty intersection of spheres. b) The derivation of reminding right side empty intersection restriction line L_β. c) The left side symmetric case, generating the third restriction Line L_α.

The edge surface separating K_{meet} from the other regions depict single point intersections of spheres, whereas elements within K_{meet} represent intersection depicting a circle with non-zero radius.

Latter conceptualization soundly amalgamates the distance among centers of the spheres with their radii. It produces a robust and general criteria to establish intersection guarantee, see Fig.14-d.

4.4 Restriction Hyperplanes

The previous derivation of the restriction lines was achieved by considering only the case involving two spheres, however, it is possible to extend these restrictions to n spheres.

Formally, this affirmation is theoretically supported by representing the n sphere radial configuration space $\mathbf{S}^n$ as the *Hilbert* space $\mathbf{C}^n$, where each dimension depicts the radius of one sphere. In element $x_{conf} \in \mathbf{S}^n$ of the n-dimensional radial configuration space can be uniquely specified by its coordinates with respect to orthonormal basis vectors

$$\hat{e}_i \in \mathbf{S}^n \mid i \in \{1, \cdots, n\} \subset \mathbf{Z},$$

which are, as expected in a *Hilbert* space, perpendicular to each other, because the radius of each sphere is independent from the others. In this manner, the previous restriction lines could be perpendicularly extruded in $n-2$ dimensions creating the restriction hyperplanes $\Phi_\alpha^{(i,j)}$.

Here again, each hyperplane divides the space in two subspaces. Configurations within the region opposite to the normal vector V_{L_α} (back of the hyperplane) represent non-intersecting spheres, see Fig.15.

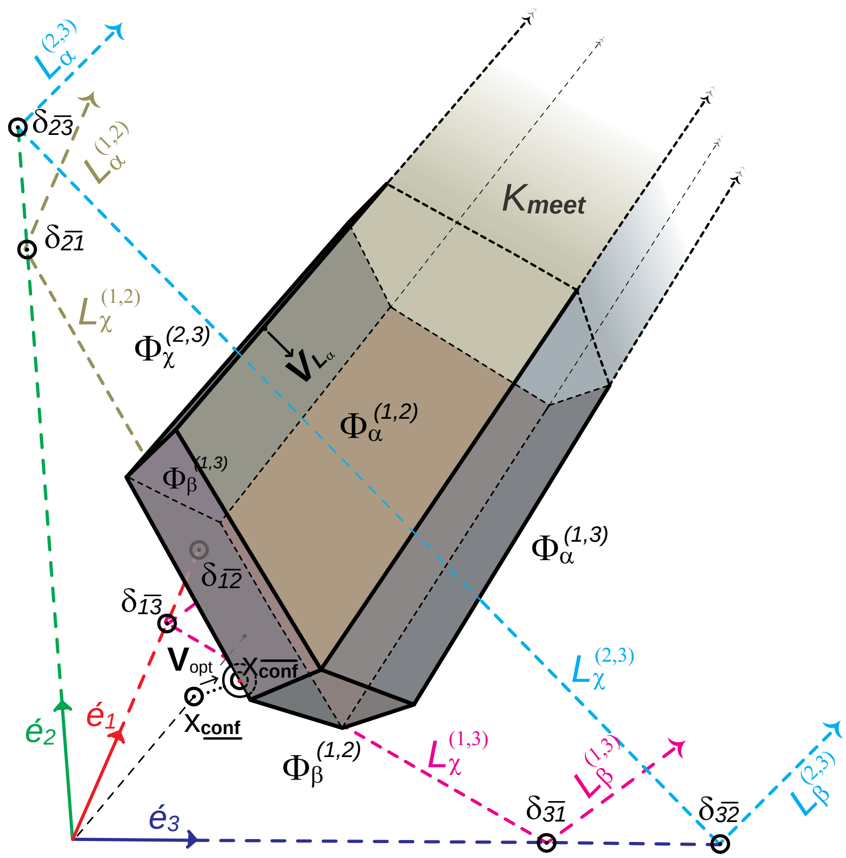

Fig. 15 The radial density space $\mathbf{Sb}^3$ containing the open polytope which delineates the subspace K_{meet}. Observe the transformation-optimization vector V_{opt} which implies an isotropic variation in the underlying density domain while creating a general dilatation within the implicit radial domain.

Even more, the set of hyperplanes expressed in their hessian form could be used to compose a matrix inequality

$$\mathbf{A}x \leq b, \tag{16}$$

where $\mathbf{A}$ is a $m \times n$ matrix, with m bounding half-spaces (normal vectors of the hyperplanes) and b represents a $m \times 1$ column vector formed by stacking the Hessen distances of the hyperplanes, i.e. an open polytope, see Fig.15.

Consider the case where $n = 3$. Three spheres implying an open polyhedron, within the radial space each line

$$L_{\alpha}^{(i,j)}\ , L_{\beta}^{(i,j)} \ \mathit{and}\ L_{\chi}^{(i,j)}$$

could be extruded in the complementary dimension creating restriction planes given by $\Phi_{\alpha}^{(i,j)}$. Next, the face cells, ridges and vertices of the polytope are found using a simple and fast implementation for *vertex enumeration* [19], see Fig.15.

At this stage, it could be conveniently established whether the current configuration is a valid, in other words, determine if the point x_{conf} belongs to the polytope. This assertion is formally given by

$$\mathbf{A}x_{conf} < b.$$

In case this assertion is held, there is no need to go through the following optimization phase because the spheres meeting on their surface, resulting the maximal density

$$\widehat{f}(\overline{x_{conf}}) = 1.$$

The opposite situations represent those degenerated configurations resulting from noise measurements and previously discussed errors. For instance, the point $x_{\underline{conf}}$ represents an invalid configuration, outside of the polytope where no intersection of spheres exist, see Fig.15.

The target solution for the latter cases necessarily implies a decay in the density, because at least one of the vector components has to be modified for the point $x_{\underline{conf}}$ in order to become a valid configuration $x_{\overline{conf}}$. This offset signifies a dilatation or relative contraction of the sphere(s) depending on the magnitude and direction of the displacement

$$x_{\overline{conf}} = x_{\underline{conf}} + V_{opt},$$

which transforms the degenerated configuration into a valid one, see Fig.15. Here, the optimal criterion to accomplish is to calculate the minimal length offset vector transformation V_{opt}

$$V_{opt} := [v_{r_1}, \ldots, v_{r_n}] \in \mathbf{S}^n,$$

retaining as much density as possible by eluding degradation of the spheres, reducing the radial variance within Eq.12.

The geometric intuitive way of finding such a vector is to find the closest point from $\underline{x_{conf}}$ on the cells or ridges of the polytope, which could be efficiently computed by perpendicularly projecting the point $\underline{x_{conf}}$ to each hyperplane

$$x^{(i,j)}_{\underline{conf}} = \underline{x_{conf}} - (V^{(i,j)}_{\alpha} \cdot \underline{x_{conf}}) V^{(i,j)}_{\alpha}, \quad (17)$$

and selecting the closest one from those points holding the assertion given by Eq.16. Although this technique is computationally efficient and geometrically correct the outcoming solution is not the optimal. Because within this space only the absolute directed distance is considered. No contribution effects of different variances are assessed, producing non-minimal density decay.

This limitation could be vanquished by considering a *homothety* transformation $\mathbf{H}(\mathbf{S}^n)$, i.e. a variance normalization of the radial configuration space inspired by the concept behind the *Mahalanobis* [3] distance.

The spatial density function of a gaussian sphere Ω_i given by Eq.12 could be conveniently reformulated in the radial domain as

$$\widehat{f}(\Omega_i, x) = e^{-\frac{1}{2}\left(\frac{r_x}{\sigma_i} - \frac{r_i}{\sigma_i}\right)^2}, \quad (18)$$

in such a way the variance of the endowed normal distribution scales the implicit defined radius r_x and the mean radius r_i of the sphere Ω_i by the factor σ_i^{-1}. This normalization mapping could be generalized for the whole radial configuration space $\mathbf{S}^n$ as

$$\mathbf{H} = diag\left[\sigma_1^{-1}, \ldots, \sigma_n^{-1}\right]. \quad (19)$$

This matrix actually represents the inverse covariance matrix Σ^{-1} of the total density function given by Eq.20. This could be easily visualized by the alternative expression[10]

$$\widehat{f}_t(x) = e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{||x - X_i||}{\sigma_i} - \frac{r_i}{\sigma_i}\right)^2}. \quad (20)$$

Based on Eq.20 and taking into account the uncorrelated radial distributions, it is clear that the underlying covariance matrix $\mathbf{H}^{-1} = \Sigma$ has zero elements outside its trace. Because of this fact, the proposed normalization $\mathbf{Sd}^n = \mathbf{H}(\mathbf{S}^n)$ could take place by applying the matrix $\mathbf{H}$ as an operator over the orthonormal vector bases of $\mathbf{S}^n$ as

$$\acute{e}_i = \mathbf{H}\hat{e}_i.$$

The euclidean metric within this resulting space is uniformly isomorphic with the density space. Displacements of the same length arising from the same position imply equal density decay in all directions reflecting different dilatation or contractions of those involved gaussian spheres. Note, this normalization takes place before the vertex enumeration for the polytope extraction has been computed, reflecting the

[10] By rewriting the exponent as a vector column and arranging in a standard form $x^t \Sigma^{-1} x$.

effects within the affine[11] strata while computing the optimal points in Eq.17, see Fig.15.

The application of the previous methods within the normalized radial configuration space $\mathbf{Sd}^n$ does not only ensures the optimal solution with minimal decay, but it also benefits from the available certainty provided from those spheres with smaller variance (higher reliable percepts) by introducing smaller displacements in the corresponding dimension of the displacement vector $V^d_{opt} \in \mathbf{Sd}^n$.

In other words, the spheres which have a wider variance can easily expand (or contract) their surfaces than those with smaller ones in order to obtain the highest possible density at the meeting operation.

This method delivers the optimal trade-off fusion while performing the management of the modeled uncertainty.

4.5 Duality and Uniqueness

In case the latter method has taken place in $\mathbf{Sd}^3$ (considering three spheres) obtaining the optimal configuration $x_{\overline{conf}} \in \mathbf{Sd}^3$, there is still a duality to solve while back mapping this configuration into the physical euclidean space. This issue is solved ia a straightforward way by computing the pair of points solution

$$J_{\wedge_{i=1}^{3}} = \bigwedge_{i=1}^{3} \Omega_i(\sigma_i(x_{\overline{conf}} \cdot \hat{e}_i), X_i).$$

In case both solutions lie within the valid[12] subspace a simple cross-check against the location of percepts which were not involved in previous calculations will robustly disambiguate the solution.

It is possible to obtain a unique solution by using four spheres for the optimization task, i.e. to represent the setup within $\mathbf{Sd}^4$.

In this way $x_{\overline{conf}} \in \mathbf{Sd}^4$ could be again mapped back into the physical euclidean space by means of the meet operator unveiling the position of the robot as

$$P_{\wedge_{i=1}^{4}} = \bigwedge_{i=1}^{4} \Omega_i(\sigma_i(x_{\overline{conf}} \cdot \hat{e}_i), X_i).$$

5 Conclusion

This approach solves the model-based self visual localization using conformal geometric algebra and gaussian spheres. The proposed method translates the statistical

---

[11] In the Hessian normal form of the hyperplanes.

[12] Above the floor and inside the modeled space.

optimization problem of finding the maximal density location for the robot into a radial normalized density space $\mathbf{Sd}^n$ which allows a very convenient description of the problem. Within this domain it is not only possible to draw the geometric restrictions which ensure the intersection of spheres, but it also attains the optimal fusion and trade-off of the available information provided from the percepts by incorporating the available information of each landmark according to its uncertainty.

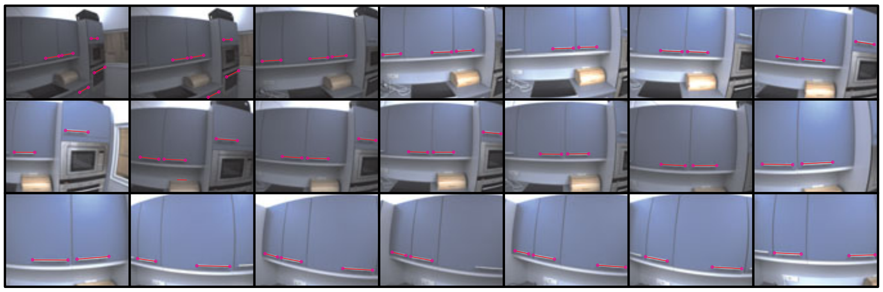

Fig. 16 The scanning sequence executing the self-localization, the results from the door-handle recognition component are highlighted.

The considered world-model[13] of the kitchen consists of 611 rectangular prisms, 124 cylinders, 18 general polyhedra with 846 faces, all arranged by 1,524 general transformations (rotation, translation and scaling) with a total of 13,853 vertices and 25,628 normal vectors composed in the scene-graph from the CAD model and verified against real furniture with laser devices, see Fig.1-a.

The global self-localization of the humanoid robot ARMAR-III [1] within the modeled environment was successfully performed using this approach. The scanning strategy takes 15-20 seconds processing 20 real stereo images. The Graph model pruning takes 100-150 ms. The hypotheses generation-validation takes 200-500 ms. Finally, the vertex enumeration takes approximately $15-50$ ms depending on the configuration.

6 Acknowledgements

The work described in this paper was partially conducted within the German Humanoid Research project SFB588 funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft) and the EU Cognitive Systems project PACO-PLUS (FP6-027657) funded by the European Commission.

[13] Human-centered environment [20].

References

1. T. Asfour, K. Regenstein, P. Azad, J. Schroder,A. Bierbaum, N. Vahrenkamp, and R.Dillmann: ARMAR-3: An Integrated Humanoid Platform for Sensory-Motor Control. In: 6th International Conference on Humanoid Robots, IEEE-RAS. 4-6 Dec. 2006, 169-175.
2. T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann: The Karlsruhe Humanoid Head. In IEEE-RAS International Conference on Humanoid Robots, 2008.
3. H. John and S. Horst: Handbook of Mathematics and Computational Science, 2006 ISBN: 978-0-387-94746-4.
4. E. Bayro and G. Sobczyk: Geometric Algebra with Applications in Science and Engineering. Birkhuser 2001, ISBN: 978-0-8176-4199-3.
5. L. Dorst, D. Fontijne, S. Mann: Geometric Algebra for Computer Science, An Object-Oriented Approach to Geometry. The Morgan Kaufmann Series in Computer Graphics, San Francisco CA, USA, 2007. ISBN: 0-123-69465-5.
6. C. Harris and M.J. Stephens: A combined corner and edge detector.In Alvey Vision Conference, pages 147 152, 1988.
7. D. Lowe: Object recognition from local scale-invariant features. In: The Proceedings of the Seventh IEEE International Conference on, Sept. 1999, 1150 - 1157.
8. D. Gonzalez-Aguirre, T. Asfour, E. Bayro-Corrochano and R. Dillmann: Model-Based Visual Self-Localization using Geometry and Graphs. In: ICPR 2008. 19th International Conference on Pattern Recognition. Tampa, Florida Dic. 2008.
9. S. Ullman: High-Level Vision. The MIT press, Massachusetts, USA, 1996. ISBN 978-0-262-71007-7.
10. S. Se, D. Lowe and J. Little: Vision-based global localization and mapping for mobile robots: IEEE Transactions on Robotics and Automation, June 2005, 364-375.
11. Duda, R. and Hart, P. (1973). Pattern Classification and Scene Analysis. John Wiley and Sons. ISBN 0-471-22361-1.
12. D.Comaniciu and P. Meer: Mean shift: a robust approach toward feature space analysis. In:Pattern Analysis and Machine Intelligence, IEEE Transactions on, May 2002, 603-619.
13. D. Gonzalez-Aguirre and E. Bayro-Corrochano: A Geometric Approach for an Intuitive Perception System of Humanoids. In: International Conference on Intelligent Autonomous Systems, Proceedings IAS-9, March 2006.
14. M.Pollefeys, L. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops and R. Koch: Visual Modeling with a Hand-Held Camera. In: International Journal of Computer Vision, January 2004, 207-232.
15. R. Hartley and A. Zisserman: Multiple View geometry in computer Vision. Cambridge University Press (2004). ISBN 0521540518.
16. O. Kallenberg: Fundations of Modern Probability. New York: Springer-Verlag, 1997, ISBN 0387953132.
17. M. Pollefeys, L. Van Goo, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops and R. Koch: Visual Modeling with a Hand-Held Camera. In: International Journal in Computer Vision. Hingham, MA, USA, 207–232 (2004)
18. T. Korn and A. Granino, Mathematical handbook for Scientists and Engineeres. Dover Publications, Dover UK. ISBN 0-486-41147-8.
19. D. Avis and K. Fukuda: A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. In: International Journal of Discrete Computational Geometry, 8:295-313, 1992.
20. T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke and J. Schröder and R. Dillmann: Toward humanoid manipulation in human-centred environments. In: Journal of Robotic Autonomous Systems. North-Holland Publishing Co. 2008 Amsterdam.