AUTOMATED ATTITUDE ESTIMATION FROM ISAR IMAGES

Marcos Avilés [(1)], Gerard Margarit [(2)], Marco Canetri[(3)], Stijn Lemmens[(4)]

[(1)] GMV Aerospace and Defence, S.A.U., Spain, Email:maaviles@gmv.com
[(2)] GMV Aerospace and Defence, S.A.U., Spain, Email: gmargarit@gmv.com
[(3)] GMV Aerospace and Defence, S.A.U., Spain, Email: marco.canetri@gmv.com
[(4)] European Space Agency (ESA), Germany, Email: stijn.lemmens@esa.int

ABSTRACT

Determining the kinematic state of objects in space is a topic of major concern for both researchers and spacecraft operators. The scientists and engineers are, e.g., interested in the influence of the attitude changes on the orbit of the object, be it for long-term propagations of the state vector or for re-entry predictions, driven by the varying geometric cross-section. For the operators the capability becomes highly important in the case of contingency situations, when communications with the satellite might be lost and solutions have to be found.

Different approaches are currently being explored, such as laser ranging, light-curves or Inverse Synthetic Aperture Radar (ISAR) techniques. Our work focuses on the latter, in which the apparent motion of the object with respect to a single radar station is used to determine the geometry and motion of the reflecting object.

This paper presents the analysis and results of applying computer vision techniques to estimate the pose of a space object only from ISAR images.

1 INTRODUCTION

The ultimate goal of this study was to achieve automated attitude (both state and evolution) extraction of space objects observed by ground radars. This will facilitate the interpretation of the observations done on non-controlled objects such as defunct satellites or rocket bodies (or parts thereof), whose kinematic state is not known. In general, the attitude evolution of a decommissioned object is supposed to be irregular at first, and then regularise slowly under the influence of external torques, depending on the inertia tensor of the object.

Knowing the kinematic state of these objects is important since on one hand it provides more information on the actual status of the observed objects and possibly leads to the analysis of the root problem cause, and on the other hand it may help designing any potential active removal mission.

One of the challenges regarding ISAR imagery is that the image plane lies quite differently from the optical case, for which computer vision techniques were usually designed. The line of sight for ISAR images is embedded in the image plane and not orthogonal to it as in optics, whereas the other dimension of the image plane depends on the rotational motion of the object. This has a strong impact in the motion estimation stage, which has to be adapted to cope with this particularity. Furthermore, ISAR images suffer from noise during the generation process, especially due to the existence of multipath reflections, and as such, algorithms have to exhibit a certain robustness with respect this kind of noise (whose multiplicative nature is also different from the additive one found in optical imagery).

Availability of ISAR images is small and acquisition campaigns are also costly. In line with these restrictions, three different scenarios were considered:

- Coarse pose estimation, based on a single image (or eventually a sequence) and knowledge (in the form of a simplified CAD model) of the target.
- Pose estimation refinement. This can be either the refinement of a single frame coarse estimation, either extracted automatically or with the assistance of an operator, or the processing of a whole sequence, where the output of one frame is considered as the coarse estimate of the following one.
- Model-less pose estimation, when there is no information about the model, but a sequence of consecutive ISAR images is available

2 COARSE POSE ESTIMATION

The basic idea for the coarse pose estimation procedure is to first acquire a training set of images or views (called templates) of the target in many different poses. Then, at runtime, a similarity measure between the input image and the templates is computed. The template with the highest score is selected as the best match, which then enables us to retrieve the pose of the camera with respect to the object.

Rather than directly comparing images, more robust features are extracted from both the input image and the training dataset and the comparison is performed using those. The choice of which features to use is not only marked by their discriminative capabilities, but also by the ability of extract them in the available training images. Focusing on the case of ISAR imagery, we are

highly restricted in the number of such training images, where typically few or no images with known ground-truth are available. Therefore, we must restrict the training dataset to synthetically data simulated through tools such as GRECOSAR [7], which simulates the entire process of transforming Doppler-Range measurements, or MOWA (Models on Orbit With an Attitude) [6], which shows how ideal ISAR images would be generated by a ground based radar.

The simulation of real imagery with GRECOSAR (in the sense that the output of the simulation is the same as the real image) can only be achieved by a deep knowledge of the model, including not only geometry but also material properties, which seems difficult in a normal operating scenario. We focus on features which can be extracted from the input images and matched against corresponding features extracted in the (less realistic) simulated imagery produced by MOWA.

The training set is thus made by sampling the whole space of viewpoints and generating reference images at each of these viewpoints (see Fig. 1).

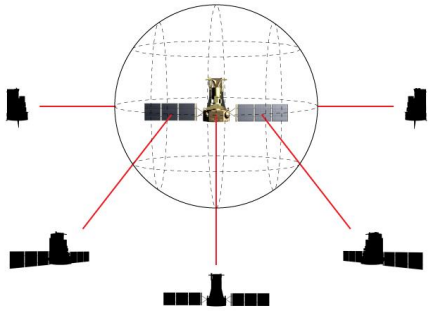

Figure 1. Sampling the viewing sphere from a discrete number of points.

Taking into account these considerations, we opted for the shape or silhouette as the method to compare the templates with the reference image. Silhouette-based methods are reasonably robust against illumination or noise and they are also invariant to scaling, translation and orientation (in the image plane). Besides, the generation of a database of silhouettes of ISAR simulations is feasible even if the knowledge of the model of the target is limited. As a drawback, the silhouette is not a very discriminative characteristic of the model (especially if the model has symmetries).

First, the reference images, generated with MOWA are processed to extract the silhouette of the target. Both for efficiency and for more robust results, these silhouettes are grouped into clusters so that the comparison between the silhouette of the ISAR image and the silhouettes of the reference images is performed hierarchically, by first comparing against the cluster representatives and then against the children of the most similar representative. Clustering is not performed spatially but based on the similarity of the reference views. Therefore, the similarity between all views needs to be computed.

To search for the best match, the silhouette of the ISAR image is also extracted and then compared against the (hierarchical) set of reference images.

Hence, there are four different tasks in the procedure:

- Silhouette extraction, both for the reference images and the ISAR images
- Computation of the similarities between the silhouettes of the reference images.
- Clustering of the reference images based on their similarities.
- Compare the silhouette of the ISAR image against the set of reference silhouettes and search for the closest match. The viewpoint used to generate that reference view will indicate the attitude of the target.

2.1 Silhouette Extraction

We evaluated a number of techniques for the segmentation of the ISAR images and the extraction of the target silhouette, and found that the differences in the results between them was relatively small.

In our particular scenario, with a clear distinction between the background and the foreground, more complex techniques do not seem to guarantee better results than the simpler ones, at the cost of being more difficult to tune (with a wrong tuning resulting in worse results, whereas simpler approaches provide a more consistent result independently of the tuning).

Furthermore, since the choice of the metric for comparing silhouettes turned out to be of more importance in the comparison between the input ISAR image and the template images, we opted for a more simple strategy of global thresholding, either using Otsu's method (which provides a threshold without any human intervention) or by setting the value manually by an operator.

In order to obtain more compact silhouettes and remove some isolated pixels or fill small holes, different morphological operations were also applied.

Fig. 2 shows some examples of the silhouette extraction procedure. The upper row presents the input ISAR images whereas the lower row depicts the resulting silhouettes.

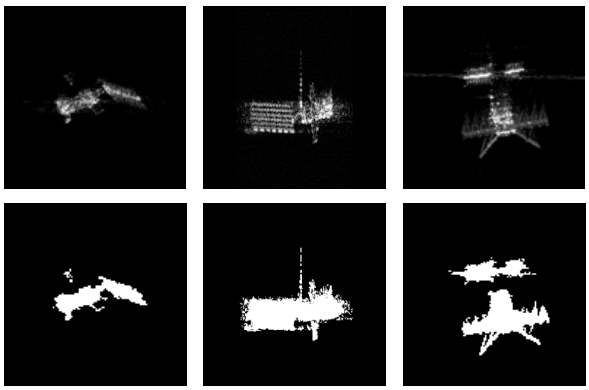

Figure 2. Examples of the silhouette extraction step.

2.2 View Clustering

As stated before, the goal of the coarse pose estimation procedure is to compare the input ISAR image against the set of reference images and to find the matching pose as the image maximizing the similarity measure. Since an exhaustive search can be computationally costly, it is better to group similar views into clusters and to perform a hierarchical search.

One might assume that reference images obtained by sampling the viewing sphere from near regions should result into similar images and therefore choosing as representative for each cluster the viewpoint located in the centre of the region. However, experimental analyses demonstrated this initial hypothesis of close viewpoints resulting in similar images does not hold in many cases. Areas of similar values are found locally, but there are also large differences between close viewpoints. Therefore, rather just grouping by distance in space, it is better to cluster viewpoints according to its own similarity.

For this task, we used the Affinity Propagation (AP) algorithm [4]. AP is a clustering method that has shown state of the art performance for a variety of unsupervised clustering tasks. Furthermore, and unlike other clustering algorithms such as *k*-means or *k*-medoids, AP does not require the number of clusters to be determined or estimated before running the algorithm.

Fig. 3 shows the results of applying the Affinity Propagation method to a set of reference images from ENVISAT. The colour code given to each combination of elevation and azimuth represents the cluster they belong to. Note that similar viewpoints tend to be clustered together (as we initially assumed) but not in a rigid manner, as it would be in case we would simply cluster views according to a spatial structure.

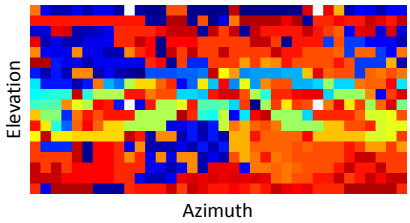

Figure 3. Clustering of reference images. The colour code represents the assigned cluster for each combination of azimuth and elevation

2.3 Shape Matching

After the silhouette is extracted, both from the ISAR image and the reference image, a metric to compare and match them needs to be defined. From the different methods evaluated, we opted for the Shape Context Descriptor [1].

Given a set of points, the shape context captures the relative distribution of points in the plane relative to each point on the shape and encodes it as a histogram. A descriptor is then formed by flattening and concatenating the histograms for all points of the shape.

An attractive characteristic of the shape context is the invariance to common deformations. Invariance to translation is intrinsic to the shape context definition since everything is measured with respect to points on the shape. To achieve scale invariance all radial distances are normalized by the median distance between all N^2 point pairs in the shape. Angles at each point are measured relative to the direction of the tangent at that point to provide invariance to rotation.

2.3.1 Rotation in the Image Plane

The distance metric just described provides scale, translation and rotation invariance. This is a powerful feature as it guarantees matching only based only on how the model is projected, without taking into account the location in the image, the scale or the (in-plane) rotation. However, once the best match is obtained, it is necessary to perform an alignment of the reference image and the ISAR image to compute the remaining degree of freedom that defines the pose of the object.

For this step, we opted for the Iterative Closest Point (ICP) algorithm [2]. ICP aims to find the transformation between two point clouds, by minimizing the square errors between corresponding entities. The algorithm iteratively revises the transformation (combination of translation and rotation) needed to minimize the distance from the source to the reference point cloud.

3 POSE ESTIMATION REFINEMENT

Once a given coarse pose estimate is obtained, either by using an automatic procedure as the one described in previous section, or with the help of a human operator,

the next step is to produce a refined estimate by minimizing a distance function around the initial estimate.

This approach can also be employed to compute the pose in a sequence of ISAR images as long as the change of orientation of the target is reasonably small. In this case, the refined pose obtained for one frame is given as coarse estimate for the following frame.

The distance function to be minimized uses the same shape metric utilized to get the coarse estimate (in the case of having been estimated automatically).

The most simple and naïve approach of compute a finer estimate of the pose starting from a coarser value is to exhaustively compute the distance between the query image and a reference image generated at every single combination of values (sampled at a finer step than the one used to get the coarser estimate) in a region of values around the starting value, and keep the solution than results in a minimum distance to the query image. Naturally, this method can be computationally very expensive if the resolution at which we want the refined estimate is high.

When the search space is quite large, simulated annealing is an alternate solution for the minimization. Simulated annealing [3][5] is a method for solving unconstrained and bound-constrained optimization problems. The method models the physical process of heating a material and then slowly lowering the temperature to decrease defects, thus minimizing the system energy.

At each iteration of the simulated annealing algorithm, a new point is randomly generated. The distance of the new point from the current point, or the extent of the search, is based on a probability distribution with a scale proportional to the temperature. The algorithm accepts all new points that lower the objective, but also, with a certain probability, points that raise the objective. By accepting points that raise the objective, the algorithm tries to avoid being trapped in local minima, and is able to explore globally for more possible solutions. An annealing schedule is selected to systematically decrease the temperature as the algorithm proceeds. As the temperature decreases, the algorithm reduces the extent of its search to converge to a minimum.

4 MODEL-LESS POSE ESTIMATION

When there is no knowledge (at least in the form of a CAD model) of the object being observed, but a complete sequence of ISAR images is available, the goal is to find correspondences between images and using those to infer at the same time the structure (shape) and the motion. This is called *Structure from Motion*.

Structure from Motion is traditionally separated in two steps. First, point-to-point correspondences are established among different views of the same scene, using assumptions and constraints on its photometry. Then, these correspondences are used to infer the geometry of the scene and the camera motion (see Fig. 4).

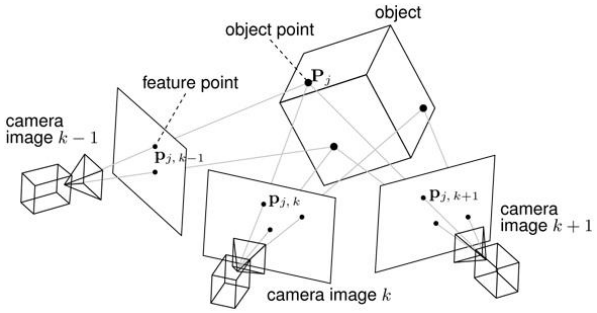

Figure 4. From point observations and internal knowledge of the camera parameters, the 3D structure of the scene is computed from the estimated motion of the camera.

4.1 Feature Detection and Matching

Feature detection concerns the automatic extraction of sparse point features from a general scene, whereas feature matching involves tracking them across a set of successive image frames.

In the context of motion estimation, features refer to point-like entities in an image, which locally have a two dimensional structure. Once features have been detected, a local image patch around each feature can be defined and a corresponding representation can be extracted from it. If those areas are similar according to some criteria, both features corresponding to different instants of time are matched. The outcome of this procedure is known as a feature descriptor or feature vector.

Different combinations of detector and descriptors were evaluated with different parameters to better understand their capabilities (number of matches, number of true matches vs number of wrong matches). The algorithm that consistently demonstrated a higher number of good matches and a higher ratio of true vs false matches was the covariant region detector combined with the SIFT descriptor.

Fig. 5 shows an example of the detection and matching of features between two consecutive frames of a sequence.

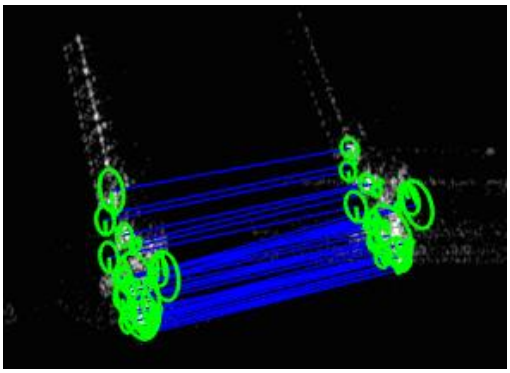

Figure 5. Example of feature detection and matching between consecutive frames.

4.2 Motion and Structure Determination

The determination of both the motion and the structure based on the sets of correspondences was made using the *factorization method*.

The factorization method was first introduced by Tomasi and Kanade [9][10] for the orthographic case and later extended for weak [11] and para-perspective models [8]. It relies on the mathematical possibility of decomposing a set of image measurements into the product of two separate factors:

$$image\ sequence \Leftrightarrow motion \times shape$$

An ISAR image lies between the purely orthographic case and the weak-perspective. Most of the mathematics are common to the first case with some modifications in the formulation.

Intuitively the projected images are considered to result from two factors: the relative *motion* between the camera and the object and the object *shape*. These are composed in a bilinear form such that if either motion or shape is constant, then the image sequence will be a linear function of the other. The motion parameters refer to all of those parameters describing the interaction between the camera and the object; namely the relative orientation and translation of the object and intrinsic camera calibration parameters. These parameters may vary from image to image in the sequence, but are the same for all features in a single image. The shape parameters describe the 3D geometric characteristics of the object and are assumed to remain constant over the sequence. Typically the 3D coordinates of features on the surface of the object are used to specify shape.

The solution is determined up to a rotation, since only the position of the world reference system has been imposed. One can fix its orientation by representing the different coordinate systems relative to that of the first frame.

The factorization method benefits from having a large set of features tracks which cover a long period of time, both of which are typically incompatible (as the more time we consider, the less likely it is to track so many features), especially for ISAR sequences, where features are more difficult to be matched.

The solution was to use a sliding window over a number of frames and apply the factorization method on each of these observation windows. Unfortunately, given that the solution for each window has an arbitrary reference transform, results cannot be directly concatenated. Nevertheless, the angular velocity independently of this arbitrary transform must be constant (at least for the period of time of the sequence). Therefore, if we can express the delta rotation between frames using an axis-angle representation, the axis should be constant within each observation window and the angle should be constant for all windows. Consequently, we are capable of extracting the angular rate, which is still quite a valuable information, but also correlated with the imposed rate required to resolve the images from the received Doppler-shifted scatters.

5 RESULTS

5.1 Manual Fitting

In order to assess the results of the different pose estimation methods, a reference attitude is needed for each image. For most of our test sequences, no telemetry or attitude information was available

For this reason, we developed a tool for manually fitting a 3D model to the ISAR image under the consideration that a human operator might perform a better estimation than an automated one. This manual procedure also allowed us to assess up to which point this human-based estimation can be used as a reference.

Fig. 6 shows an example of the manual fitting of the ENVISAT model to one frame of the sequence. One important point to note is that the intrinsic rotation used to produce the ISAR images in this sequence was not correct and therefore the cross-range scaling was also incorrect. This translates into an aspect ratio different from the reality. In order to achieve a more accurate fitting, the aspect ratio of the image need to be adjusted at the same time the model was aligned was the actual image.

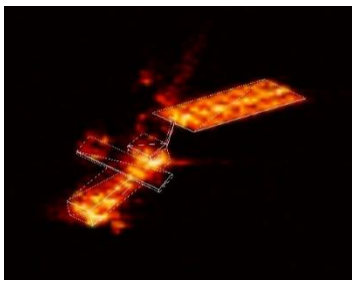

Figure 6. Manual fitting of a frame from sequence Envisat_12d117. Aspect ratio needed to be corrected (approximately in 30%) for a more accurate fit.

The main consequence is that a new degree of freedom is introduced (and thus more uncertainty). The operator needs not only to determine the pose of the target but also the correct aspect ratio that produces a better alignment.

Additionally we also studied the repeatability of the manual fitting, that is, how close the estimates obtained by the same trained operator are if he analyses the same sequence several times. Our approach was to repeat the fitting procedure three times starting from scratch (not relying in a previous fit) and compare the three different estimates.

5.1.1 Envisat_12d117

Fig. 7 depicts the attitude estimation (expressed as yaw, pitch and roll) for these three different tests. Note that we focused in the second part of the Envisat_12d117 sequence, where ENVISAT is seen from a more oblique view, and better fittings are expected.

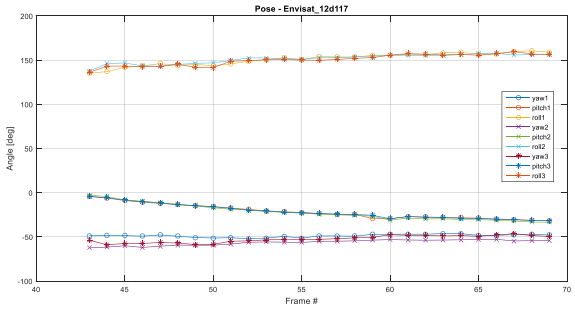

Figure 7. Manual attitude estimation for sequence Envisat_12d117.

At a first look we can see that similar attitudes were obtained for the three cases, which is reasonable as the same operator was responsible for them (using different operators would introduce more uncertainty as you then compare the spatial abilities of one operator against those of other one). However, if we look into more detail, especially in the yaw component we find a more substantial difference. Fig. 8 shows the same results but only focusing on the yaw estimation. Here we can clearly see that differences up to 15 degrees can be found between two different experiments.

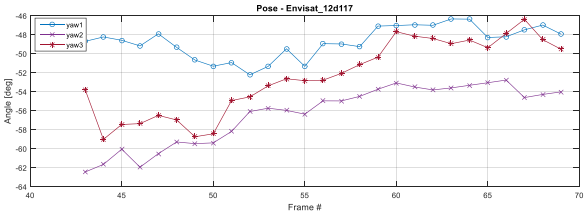

Figure 8. Manual attitude estimation for sequence Envisat_12d117. Only yaw.

Still, even if there are some significant differences we can identify a similar evolution in all the cases. To verify this hypothesis we computed the angular rate out of the three experiments and confirmed that they are all quite similar. Tab. 1 presents the statistics of the estimated angular rates.

Table 1. Statistics of the manually estimated angular rate (in deg/s) for sequence Envisat_12d117

Test #	Mean	Std. Dev.]	Median
1	0.242469	0.1456573	0.238304
2	0.233344	0.196166	0.154032
3	0.254115	0.227907	0.173437

We can conclude that the fitting of the first frame acts as a bias, derived from the fact that seven degrees of freedom (six from the pose plus one from the image aspect ratio) have to be estimated in this initial frame, but in the following frames, where only three (just the change in rotation of the target) need to be determined, the operator produces more consistent outputs. The smaller the resolution of the images is, the larger this bias will be (as it is the case for sequence Envisat_12d117) because there will be a wider range of poses that could fit to the same low resolution ISAR image.

5.1.2 Envisat_13d317

Tab. 2 presents the evolution and the statistics of the estimated angular rate for the sequence Envisat_13d317. The standard deviation is a little bit high, but this is due to the configuration of ENVISAT along the sequence, which is seen mostly from a nadir point of view and therefore there is more uncertainty in the estimation of one of the components.

Table 2. Statistics of the manually estimated angular rate (in deg/s) for sequence Envisat_13d317

Mean	Std. Dev.]	Median
3.881096	1.199530	3.681769

5.2 Coarse Pose Estimation

Different ENVISAT models were evaluated trying to produce MOWA simulations closer to the actual ISAR images in the sequences. Through visual inspections one could notice a clear difference on how the solar panel appears in each sequence, especially if we consider the silhouette rather than the actual reflectivity. Whereas for sequence Envisat_12d117, it is can be seen as a compact solid plane, for Envisat_13d317, it appears as a set of high reflectivity points (corresponding to the attachment points where the panel was stowed) together with a line (corresponding to the cabling).

The model shown in Fig. 9 (left), was selected for sequence Envisat_12d117 and whereas the model in Fig. 9 (right), was selected for sequence Envisat_13d317.

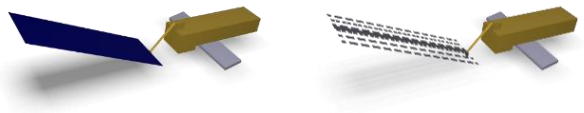

Figure 9. ENVISAT models. Left, envisatAC_noant; right, envisatAC_apSa_1cab_large_noant.

5.2.1 Envisat_12d117

Fig. 10 summarizes the results of the coarse pose estimation procedure for one part of the sequence Envisat_12d117 (the whole set of results has not being included in this report for brevity, but can be made available under request).

The figure consists in a matrix of images, where each row corresponds to each frame of the sequence. The first column depicts the silhouette of the actual ISAR image (named query image) for which the pose is to be determined. The following columns (from second to eleventh) correspond to the silhouettes of the best 10 matches in increasing order of distance (that is, the best match is the second column). In each of these cells and below the silhouette of the reference image there are two rows of numbers. The upper row shows the distance between the query image and the reference image, whereas the bottom row gives the orientation of the viewpoint (as azimuth, elevation and rotation in the image plane).

It is important to remark than the shape matching is performed using the azimuth and elevation (that is why the value is an integer, as it corresponds to one of the viewpoints of the sampled viewing sphere). Therefore two silhouettes will still be a good match even if one is a rotated version of the other. It is the estimated rotation in the image plane what brings them into alignment. Unfortunately, this estimation does not always provide the correct solution. The alignment method consists in a function minimization that is dependent on the initial point and therefore it might get stuck in a local minimum.

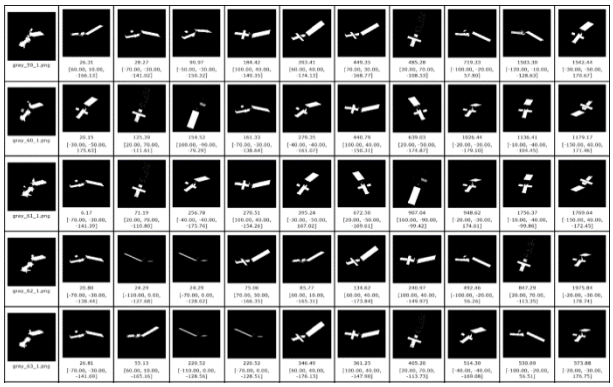

Figure 10. Best 10 solutions for frames 59 to 63 of sequence Envisat_12d117

Exploiting Temporal Coherence

As already seen the most likely match (or, at least, the one an operator would choose as the most likely one) is not necessarily the one with the best score. Nevertheless, it is usually found between the best *N* scoring possibilities (with *N* being typically 5-10).

Although the proposed solutions are reasonably similar to the input image, they correspond to quite different configurations. Without any additional knowledge, it is not possible to automatically determine which of those correspond to the reality (in some of the cases, not even for a human is possible). However, if rather than just one ISAR image, we have a sequence of consecutive images, we can introduce extra restrictions and reduce the set possibilities to those that provide a more coherent movement. This means that the target orientation chosen at a given frame should be similar to the one chosen in the previous frame, and similarly, the choice for the next frame should be close to the current one.

This problem can be considered a graph optimization procedure where the goal is to select the combination of solutions that result in a minimum rotation of the target and taking into account that the inter frame rotation should be as close as possible to a known value (typically the intrinsic rotation used to derive the ISAR image).

The optimization only considers azimuth and elevation, and not the camera angle, as the latter does not give consistent results in all the cases. In fact, this inconsistency translates into errors of 180 degrees (a whole rotation of the model in the image plane) which would highly penalize a proper solution in terms of azimuth and elevation.

Figure 11. Best 10 solutions for frames 59 to 69 of sequence Envisat_12d117 after constraining the search space.

Fig. 11 shows the sequence of solutions that minimize the rotation of the target for frames 59 to 69 (for better readability, the figure only presents a smaller set of images than those considered in the optimization). Based on this solution we can now estimate the attitude and angular rate for the sequence of frames.

Fig. 12 shows the evolution of the target attitude for frames 43 and 69. Note that we are computing a coarse estimate, where two of the angles are computed with a 10 degrees sampling step (only the third angle is computed by an optimization method with floating point accuracy). The strange behaviour at frame 64 is due to an incorrect determination of the rotation in the image plane. Remember the matching is performed only taking into consideration the azimuth and elevation of the camera and the last angle is determined by aligning the 2D projections, which leads to incorrect alignments sometimes (for elongated shapes, this error is usually 180 degrees, corresponding to an alignment with a complete rotation of the model in the image plane).

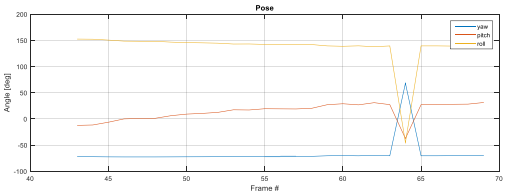

Figure 12. Coarse attitude estimation for sequence Envisat_12d117, frames 43 to 69.

Tab. 3 shows the statistics of the angular rate having removed the outlier at frame 64. The values are close to the manual estimate, but the standard deviation is large and comparable to the mean value (still we have to remember that this is based on a reasonably coarse sampling step)

Table 3. Statistics of the estimated coarse angular rate (in deg/s) for sequence Envisat_12d117, frames 43-69.

Mean	Std. Dev.]	Median
0.2683153	0.245363	0.183042

5.2.2 Envisat_13d317

When attempting to exploit the temporal coherence of the images, we first observed that due to orientation of ENVISAT during this sequence, which is seen mostly nadir, there are many frames with strong specular reflections that translate into noisy images where a suitable silhouette of ENVISAT can hardly be obtained

The results of the algorithm for these particular images do not a have any similarity with the results obtained in temporally adjacent frames, which prevents exploiting fully temporal coherence to constraint the results to one solution. Furthermore, these phenomenon is cyclic and happens every 10 images. Therefore, we can only try exploiting the temporal coherence for subsequences of up to 10 frames.

Fig. 13, red line, shows the set of solutions that minimize the rotation of the target for the subsequence going from frames 9 to 19. As we can observe, although this set of solutions guarantee a minimum rotation, they are not the mostly likely set (which could be the one depicted in blue).

Figure 13. Best solutions for frames 9 to 19 of sequence Envisat_13d317 after constraining the search space. Red line, solution with minimum distance; blue line, most likely solution.

The problem we face is that the silhouettes extracted from the ISAR sequences and those obtained from the MOWA simulations cannot be matched univocally (in many cases, even for a human it would be difficult to say which reference silhouette corresponds to the silhouette extracted from a given ISAR image). Even if we see that the mostly likely reference silhouette is found between the best scoring solutions, we cannot guarantee that others wrong reference views are also included among this set of best scoring possibilities. Furthermore, and we saw it in this case, the same erroneous view is included in all the solutions of the sequence.

5.3 Pose Estimation Refinement

Once a given coarse pose estimate is obtained, either by using an automatic procedure as the one described in previous section, or with the help of a human operator, the next step is to produce a refined estimate by minimizing a distance function around the initial estimate.

This approach can also be employed to compute the pose in a sequence of ISAR images as long as the change of orientation of the target is reasonably small. In this case, the refined pose obtained for one frame is given as coarse estimate for the following frame.

5.3.1 Envisat_12d117

We focus our analysis in the second part of the trajectory, where the solar panel is seen completely, and better conclusions can be obtained

Stand-Alone Images

We first analyse the refinement of the initial pose given by the coarse pose estimation algorithm but taking as input the result of this method and then refining each image independently (temporal coherence was only used to constraint the search space of the coarse pose estimation). As the sampling space used for the coarse estimation was 10 degrees, we restricted the optimization search space to ±5 degrees (in both elevation and azimuth) around the initial coarse estimate. Besides, given that the search space is small, we opted for exhaustive search as the optimization method (simulated annealing would make sense only if the search space is rather large).

Fig. 14 shows the estimated attitude of ENVISAT for the last part of the sequence. The jumps in the attitude corresponds to incorrect estimates of the rotation in the image plane.

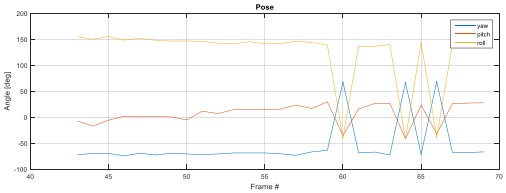

Figure 14. Refined attitude estimation for sequence Envisat_12d117, frames 43-69

Fig. 14 and Tab. 4 show the statistics and evolution of the angular rate having removed these outliers. If we compare these values with those reported in Fig. 12 and Tab. 3, which correspond to the starting point for the refinement, we see that the angular rate estimate has worsened (if we consider the manual fitting as the reference). The reason for this is that for this sequence, where the rotation of the target is small, all the images are matched with the same reference viewpoint and the only difference between estimates was only due to the

rotation in the image plane. However, during the refinement procedure, the viewpoint is also being modified, which brings more uncertainty and higher errors.

Table 4. Statistics of the refined angular rate (in deg/s) for sequence Envisat_12d117, frames 43 to 69.

Mean	Std. Dev.]	Median
0.805023	0.482144	0.677205

Exploiting Temporal Coherence

The second experiment relies on the temporality of the sequence and applies the refinement procedure sequentially, so that the output attitude at one frame is used as input attitude for the following one. The attitude for the first frame is obtained from the coarse estimation procedure after the search space constraining (that is, the input for the first frame is the same as for the stand-alone images refinement presented in previous section).

Fig. 15 shows the estimated attitude of ENVISAT for frame 43 to 69 of the sequence. As it happened in the stand-alone images refinement, the jumps in the last part correspond to incorrect estimates of the rotation in the image plane. Compared to that experiment, we see that both pitch and roll result in similar values, but yaw has a different behaviour, which can also be a consequence of the error accumulation from one frame to the other.

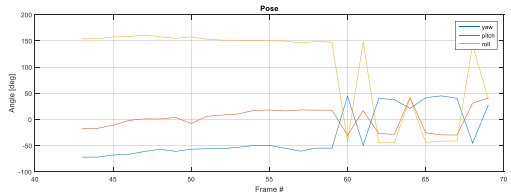

Figure 15. Refined attitude estimation for sequence Envisat_12d117, frames 43-69.

Tab. 5 shows the statistics the angular rate, but constraining the estimation to frames 43 to 59 in order to avoid the last part, which we already know it is erroneous. We now find that the mean and median are aligned, which is always a good indication for this type of problem. Standard deviation is almost the same as in previous experiment.

Table 5. Statistics of the refined angular rate (in deg/s) for sequence Envisat_12d117, frames 43 to 59.

Mean	Std. Dev.]	Median
0.673765	0.439934	0.703470

5.4 Model-Less Pose Estimation

This section presents the results of the model-less pose estimation procedure, based on Shape-from-Motion techniques

5.4.1 Envisat_12d117

Sequence Envisat_12d117 can be divided into two clear sections: during the first half, the solar panel is almost perpendicular to the line-of-sight and results in difficult estimates even for a human operator, and the second half, where ENVISAT is shown in a more oblique view facilitating the estimate of its pose.

Fig. 16 and Tab. 6 present the results of the angular rate estimation only focusing on the second half of the sequence and using a more optimized set of parameters suitable for only this part. The resulting angular rate is very close to the one obtained by a human operator and also the standard deviation is smaller, indicating the estimates are consistent.

Table 6. Statistics of the estimated angular rate (in deg/s) for the second half of sequence Envisat_12d117.

Mean	Std. Dev.]	Median
0.336426	0.185663	0.311379

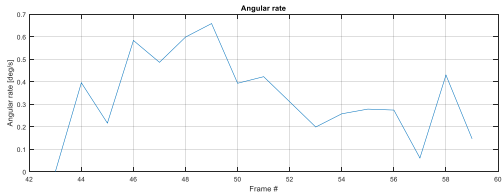

Figure 16. Evolution of the estimated angular rate for the second half of sequence Envisat_12d117.

Still it is important to note that the resolution of the images for this sequence is very small (the size of ENVISAT in the image is only around 60×60 pixels) so the accuracy will be always limited.

5.4.2 Envisat_13d317

Similarly as we proceeded for Envisat_12d117, it is advisable to analyse the sequence in sections with more homogeneous characteristics (and not covering especially noisy parts)

Tab. 7 presents the estimated angular rate for three different sections of the Envisat_13d317 sequence. We can clearly see how the estimated rate is more stable and in a similar range for three cases. Although lower than the manual fit, it is important to note that the manual estimation had a larger standard deviation of 1.19 deg/s and therefore the comparison needs to be performed judiciously.

Table 7. Statistics of the estimated angular rate (in deg/s) for the second half of sequence Envisat_12d117.

Frames	Mean	Std. Dev.]	Median
9-19	2.807013	0.587178	2.726546
37-47	2.943567	0.899692	2.637382
54-63	2.516819	0.725439	2.677517

Images in sequence Envisat_13d317 have a much larger resolution than those in Envisat_12d117 so we also analyzed the impact of the resolution on the quality of the results. For this experiment, we estimated the angular rate at different image resolutions, expressed as a percentage of the original size in each dimension (original images are all have a height of 641 pixels, so a 10% size means an image 64 pixels height).

As we already noted before, rather than analysing the sequence as a whole, it is better to focus the analysis in different sections with more homogeneous characteristics (homogenous within the section, the sections can have different properties). Fig. 17 presents the same analysis but for different sections. The left plot corresponds to frames 9 to 19 whereas the right plot corresponds to frames 37 to 47.

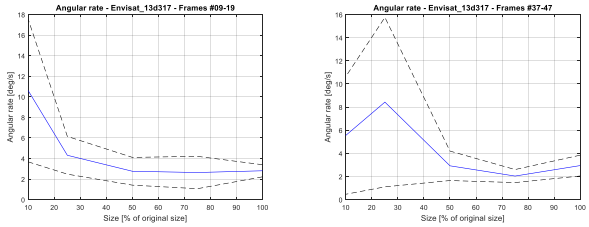

Figure 17. Estimated angular rate at different sections of sequence Envisat_13d317 and increasing resolutions; frames #9-19 (left) and frames #37-47 (right)

Given that the casuistry is rather high and that the tests at each resolution were not individually tuned (the same parameters chosen for 100% size where used for the rest of sizes), we cannot derive a definitive conclusion, but the tendency shows that up to half the size (in each dimension, so one quarter in area), which translates into a height of approximately 320 pixels, there is an stable behaviour, with a progressively increase in the standard deviation, but going lower than 50% the estimates become considerably worse.

One explanation for this effect is that the feature detection is multi-scale and therefore the same features are detected even at a smaller resolutions. The increase in the error mostly comes for the smaller precision. However, there is a point where the image becomes so small that those features are no longer detected (and/or matched), and therefore a more exponential growth of the error is experimented.

6 CONCLUSIONS

This section provides a summary of the conclusions obtained from the previous results

As a general remark, it has been shown that the quality of the ISAR images is crucial. The resolution does not only have an important impact in the accuracy but also in the robustness. Small images, such as those in sequence Envisat_12d117, where ENVISAT occupies merely 40×40 pixels should not be used in a real campaign. A reasonable cross-range scaling is also needed for silhouette based methods. Clearly, the only way of correcting the scale is by knowing the exact intrinsic rotation of the target, which is what we try to obtain, but in some of the sequences used in the activity error was considerably higher than desirable, adding an extra degree of freedom to be determined.

Silhouette-based Pose estimation

The silhouette of an object cannot unambiguously describe the shape of an object and different viewpoints can result in almost the same silhouette. Fig. 18 illustrates this problem, where we see how close the two proposed solutions are between themselves (apart from the rotation in the image plane which we have to remember that is not considered in the matching) and it is unclear which of both provides a better match, even if both correspond to a difference of 60 degrees.

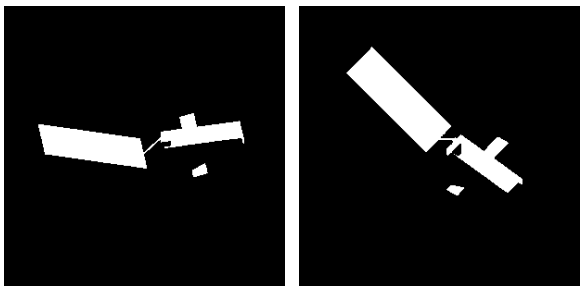

Figure 18. Two possible matches for frame 67 of sequence Envisat_12d117.

This is even more of an issue for ISAR imagery where the extraction of the clean silhouette is a problem on its own due to the noise and specular reflections that mask the shape of the target. On the other side, the generation of valid reference images to be used for the matching (and hence to extract the pose of the satellite) is a difficult task. MOWA allows real-time generation of simulations, but uses a very simplified computational model. On the contrary, GRECOSAR allows the simulation of the full ISAR process but requires a very deep knowledge of the target, both in shape and materials, to produce a simulation comparable to the reality. Whenever this knowledge is not available (which is the most often case), everything is reduced to the ability of the modeler to introduce artificial details in the model in a trial-and-error fashion so that plausible imagery is produced.

The reference set of images was produced taking into account two degrees of freedom, elevation and azimuth, as the distance metrics used for the comparison were rotation invariant (the third degree of freedom is the rotation in the image plane). This highly decreased the search space and hence the probability of false matches (for a sampling step of 10 degrees, we reduced the number of possibilities from 23328 to 648). However, the final step of determining the rotation in the image plane

turned out to be a challenge and did not guarantee coherent results. Different methods were tested but none of them could overcome the fact that the coarse shape of the silhouettes of both the ISAR image and the reference image can make the algorithm find a better fitting (in terms of distance) than the real alignment or, at least, the one an operator would determine

Additionally, we also found that the input ISAR images are not guaranteed to have a correct cross-range scaling. This was somehow expected as the only way of correcting this scaling is by knowing the exact intrinsic rotation of the target, which is what we try to obtain, but the error was considerably higher than desirable, adding an extra degree of freedom to be determined. In some cases (for instance, in the ENVISAT sequences), an operator is able to find an approximate value for this aspect ratio correction, but there are some other scenarios (like GOCE sequence), where it is also difficult for a human to determine a better fitting is obtained by changing the aspect ratio or the attitude (see Fig. 19).

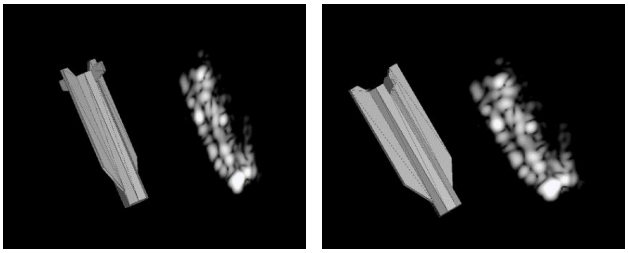

Figure 19. Two different fittings for the same ISAR image, only changing the aspect ratio.

Due to all these uncertainties, we found that the best match, according to the distance metrics, does not typically correspond to what we, as humans, would have considered as the best option. Still, this human based choice is likely to be among the first 5-10 solutions.

Including temporal coherence helps to reduce the problem and in some sequences it produces a reasonable sets of solutions. However, for other sequences, especially those were there angular rate of the target is small, the same erroneous solution is found for all frames and makes the graph optimization algorithm fail because the distance for this path of solutions is lower than that from the most feasible one.

Similarly, pose optimization, either for refining a coarsely estimated pose or for processing a sequence, tends to diverge in many cases, again due to the lack of discriminatory power of the silhouettes (considering the whole chain of limitations: the silhouette extraction from ISAR images but also the low resemblance of the reference images to the reality).

Model-Less Pose Estimation

Shape-from-Motion algorithms applied to model-less pose estimation are a promising alternative and are able to provide good results, although restricted to the estimation of the angular rate. They are, nevertheless, quite sensitive to outliers, which in the case of ISAR imagery, tend to occur much more often than with optical images. Different robust methods have been introduced to guarantee more consistent results during a whole sequence and a significant improvement was obtained, but still more reliable methods are desirable.

Detected features do not correspond with points of high reflectivity, but typically with areas with of large variation. From a human point of view, the association of high intensity points with reflections from the satellite and their discrimination from noise can be done in many cases (even if there is no concrete knowledge of the target, one can interpret the reflections), but from a machine point of view, it is not straightforward to distinguish between high intensity points due to noise or reflections from the satellite. That is why the intensity at a point is not the driving procedure to extract features, but the texture and variations around it. For ISAR imagery, these areas tend to be less stable than in optical imagery. Manual selection of large reflections corresponding to the target, but an automatic tracking of them is a path to explore in the future.

Angular rate is also a key factor for the algorithm, because at high velocities features can only be tracked for a small number of frames, which result in less accurate and stable estimates. Image size is also an important factor. For images where the target is between 500-1000 pixels, the impact of the size if mostly in the accuracy (smaller size translate into lower accuracy), as the implemented feature detection methods are scale-invariant and the same features are found. However, for smaller sizes, the impact is much larger, as those features found at larger images are no longer detected, and the ones found are less reliable, which makes the error grow (both in mean and standard deviation).

Manual Pose estimation

Manual pose estimation also has a large uncertainty, especially for small images, at least in terms of absolute estimation of the pose. The same operator will likely provide different solutions (with differences which can be up to 20 degrees) if he is given the same image without any type of initialization or guess. Nevertheless, he will probably be much more consistent in the estimation of the delta transformations. Starting from an initial pose, he will be able to estimate the change of orientation between frames if there is some temporal coherence (once he has adjusted a few frames, he will be able to "predict" how much the target will have changed in the following frame).

7 REFERENCES

1. Belongie, S., Malik, J. and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(4): 509–522.
2. Besl, P. J. and McKay, H. D. (1992), A method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(2): 239–256.
3. Černý, V. (1985), Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, *Journal of Optimization Theory and Applications*, **45**: 41–51.
4. Frey, B. J., Dueck, D. (2007), Clustering by passing messages between data points. *Science*, **315**: 972–976
5. Kirkpatrick, S., Gelatt Jr, C. D. and Vecchi, M. P. (1983), *Optimization by Simulated Annealing*, *Science*, **220** (4598): 671–680.
6. Lemmens, S. and Krag, H. (2013), Sensitivity of automated attitude determination form ISAR radar mappings, *Advanced Maui optical and space surveillance technologies conference*: 768–779.
7. Margarit, G., Mallorqui, J. J., Rius, J. M. and Sanz-Marcos, J. (2006), On the usage of GRECOSAR, an orbital polarimetric SAR simulator of complex targets, for vessel classification studies, *IEEE Transactions on. Geoscience and Remote Sensing*, **44**(12): 3517–3526.
8. Poelman, C. and Kanade, T. (1993), A Paraperspective Factorization Method for Shape and Motion Recovery, Carnegie Mellon Univ., Pittsburgh, PA, Tech Rep. CMU-CS-92-208.
9. Tomasi, C. and Kanade, T. (1991), Shape and motion from image streams: A factorization method, Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-91-105.
10. Tomasi, C. and Kanade, T. (1992), Shape and motion from image streams under orthography: A factorization method, *Int. J. Comput. Vis.*, **9**(2): 137–154.
11. Weinshall, D. and Tomasi, C. (1995). Linear and incremental acquisition of invariant shape models from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(5): 512–517.