# An approach to grasp planning based on the Natural Grasping Axis with a real 3D vision system 

Cédric Michel ${ }^{1,2}$, Véronique Perdereau ${ }^{1,3}$, and Michel Drouin ${ }^{1,4}$<br>${ }^{1}$ L.I.S.I.F.Université Pierre et Marie Curie, 4 place Jussieu 75252 Paris cedex 05 France<br>${ }^{2}$ cedric.michel@lisif.jussieu.fr<br>${ }^{3}$ veronique.perdereau@upmc.fr<br>${ }^{4}$ mid@ccr.jussieu.fr

Summary. We address here the problem of grasping 3D objects whose geometrical envelopes are provided by a visual system. A previous article described a grasp planner derived from the human grasp for virtual objects modelled with a rendering software. We made evident that, during the approaching phase, human beings shape their hands to comply with what we call the Natural Grasping Axis, parallel to the palm and surrounded by the thumb facing the other fingers. However, to estimate the grasping axes for real objects, it is crucial to extract their 3D envelope from their environment. In this paper, we propose to extend this original methodology for grasp planning by implementing an algorithm for 3D reconstruction with multicameras: the Voxel Coloring. It consists in decomposing the scene to be reconstructed into small elements of volume projected onto the image set. A complete simulation example of the approach is given to illustrate the performance of this methodology.

## 1 Introduction

Many multi-fingered grippers have been developed in the past few years in order to perform remote manipulations in hazardous environments with a greater dexterity than a simple gripper. Therefore a lot of research has been done in the field of grasping and manipulating, especially for isolated targets [1-2]. Grasp synthesis, defined as finding a set of contact points on a given object geometry, can be achieved either through optimisation or using a constructive algorithm. The addressed problem becomes very difficult in three-dimensional space. Indeed, grasp synthesis is a complex issue due to the large number of degrees-of-freedom of a hand and the variety of constraints involved. A satisfying choice of finger configurations for grasping any 3D objects with a mechanical multi-fingered hand depends on three main criteria: the grasp must be cinematically feasible, free of collisions and suitable for the task to be performed (Fig.1).


Fig. 1. Definition of the optimal grasp

As a result, the optimal grasp belongs to the intersection of the set of the possible grasps in the absence of obstacles, the set of the reachable grasps in a constrained environment and the set of the grasps adapted to the nature of the task to be executed. However, defining the set of the grasps adapted to the nature of the task requires the planner to be able to clearly identify the object in a database, probably using an expert system. At the stage of our research, it is not the purpose of this paper which searches for the optimal grasp of an object whose functionality is unknown. The original methodology we proposed to solve this complex issue is to adapt the usual human grasping behaviour to our robot by implementing an iterative algorithm which evaluates the socalled Natural Grasping Axis for each target. To perform our simulations, we previously used the rendering software 3DS in order to model a wide range of various objects whose geometry was, consequently, perfectly known. In this paper, our goal is to adapt this algorithm for real objects. It implies to extract the 3D geometry of the targets in the scene with given cameras and their respective images. We use a scene-based approach: Voxel Coloring rather than traditional methods such as stereovision. It consists in decomposing the landscape to be reconstructed into elementary cubes called voxels, projected onto the image set. Then an algorithm that passes through these voxels and colors those that belongs to a surface in the scene. This modus operandi is perfectly similar to an artist sculpting a raw block of marble. Voxel Coloring was initially developed to build photorealistic 3D photos. The discretized set of voxels representing the target object is actually perfectly adapted to calculate the N.G.A.. That is the reason why we chose this method to solve our grasping topic. The paper is organized as follows. In the next section, we introduce a brief overview of the original grasping methodology developed in case of an isolated target. Then in section 3 , we outline the algorithm used to acquire the 3D geometry of the target objects, before presenting results obtained for a small teapot. Finally, we discuss future directions concerning this issue.

## 2 The Grasping axis generation algorithm

Most of the former studies concerning grasping involve either simplified representations of the object, based on elementary geometric models [3], or neural networks selecting grasp configurations they have been trained on. These methods are particularly efficient when basic objects are involved, but show their limits in the manipulation of more complex shapes. The original methodology we developed imitates the human hand with the aim of elaborating a universal grasping strategy.

### 2.1 A concept inherited from the human grasp

Many research works have focused on biological grips [4] leading to the two generic grasp configurations enabled by a hand: the power grasp and the precision pinch. During the preliminary approach, before grasping, human beings shape their hands to match one of the grasp modes explained above. A deeper investigation demonstrates that, as a matter of fact, the human hand favours large facing surfaces of the solid. Therefore, during the preshaping phase (Fig.2), the whole geometry of the object, and especially these facing surfaces, lead to orientate the human hand along a grasping axis, parallel to its palm and surrounded by the thumb facing the other fingers : the Natural Grasping Axis.


Fig. 2. The Natural Grasping Axis for a human hand

Intuitively, the Natural Grasping Axis is a straight line as parallel as possible to the whole surface of the targeted object. We demonstrated that, considering a solid whose surface is reconstructed with small facets (like triangles for instance), this axis is the most parallel one to all couples of surfaces used for the grasping [5-6].

### 2.2 The importance of the facing surfaces

As explained before, the human hand actually favours large opposite surfaces of the solid. Thus, the N.G.A. results from the minimization of a quadratic function, representative of the couples of the opposite faces contributing to
the whole geometry of the target. Given $F_{i}$ and $F_{j}$ two elementary surfaces on the object surface, their respective normal vectors $n_{i}$ and $n_{j}$ and their inter-distance vector $p_{i j}$, we define $\alpha_{i}$ et $\alpha_{j}$ as the angles between pij and the normal vectors of the two faces. In our approach, we quantify the opposition of both surfaces $F_{i}$ and $F_{j}$ as the minimum of the angles $\alpha_{i}$ and $\alpha_{j}$ (Fig.3):

$$
\begin{equation*}
F_{i} \text { and } F_{j} \text { facing } \Leftrightarrow \alpha_{i j}=\max \left(\alpha_{i}, \alpha_{j}\right) \text { minimal } \tag{1}
\end{equation*}
$$

The minimal angle $\alpha_{i j}$ as calculated above is the critical angle $\alpha_{i j}^{c}$.


Fig. 3. The facing surfaces

Furthermore, a robotic grip is limited by the maximal gap dec between fingers. As a consequence of this key point, overly distant facing surfaces must not interfere in the estimation of the N.G.A. resulting to a new formulation of the facing surfaces from the point of view of grasping:

$$
\begin{align*}
F_{i} \text { and } F_{j} \text { facing } \Leftrightarrow \alpha_{i j} & =\max \left(\alpha_{i}, \alpha_{j}\right) \text { minimal }  \tag{2}\\
\&\left\|p_{i j}\right\| & <d_{\infty}
\end{align*}
$$

We iterate a process which associates a surface with its facing surfaces (Fig.4). It is worth noticing that the number of triangles Fi involved in the triangulated representation can explode but we reduce it by creating the socalled mega-surfaces. Those result from a criterion of parallelism for the unified triangles and their areas are adapted to the size of a finger. So, after the step of 3D acquisition, we unify small triangulated facets into a set of larger surfaces whose dimensions are finger sized.


Fig. 4. Example of facing surfaces calculation

### 2.3 The quadratic criterion

Once the facing couples have been calculated, we express a quadratic criterion whose minimization leads to the Natural Grasping Axis. It is founded on the quadratic sum of the angles between the normal vectors of the opposite surfaces and the Natural Grasping Axis. Considering only one couple of facing surfaces $\left(F_{i}, F_{j}\right)$ associated with a critical angle $\alpha_{i j}^{c}$ (Fig.5) the local grasping axis becomes a local grasping plane. As a matter of fact every axis in this specific plane is a suitable grasping axis with only one couple of facing surfaces.


Fig. 5. Local definition of the criterion

To determine this plane called $P_{i j}$, we have to maximize the cosines of the angles $\beta_{i}$ and $\beta_{j}$ between $P_{i j}$ and the normal vector $n_{i}$ and $n_{i}$, respectively which is equivalent to finding a plane $P_{i j}$ as perpendicular as possible to these normal vectors. Each couple of facing surfaces contributes to a possible direction oriented by $u$ for the Natural Grasping Axis. That is why we add the contributions of all the considered couples in the following quadratic criterion $J(u)$ explained in [5] with the intention of getting a unique global straight line oriented by the vector $u^{*}$ which minimizes this criterion:

$$
\begin{gather*}
J(u)=\sum_{i=1}^{N_{f}} \zeta_{i E_{i}} \frac{\sum_{j \in E_{i}}^{N_{i}} S_{j} \cos \left(\frac{n_{j}-n_{i}}{\left\|n_{j}-n_{i}\right\|^{2}}\right)^{2}}{N_{i}}  \tag{3}\\
\forall i, \forall j \in E_{i} /\left(F_{i}, F j\right) \text { opposite \& } u^{T} u=1
\end{gather*}
$$

In the previous expression, $i \in\left[1, N_{F}\right]$ represents the index related to the $N_{F}$ surfaces of the solid. Each surface $F_{i}$ of area $S_{i}$ matches one or more facing surfaces $F_{j}$ of area $S_{j}$ (we call $N_{i}$ their number). These areas enable to favor large surfaces in order to ensure the grasping stability. Consequently, couples of surfaces with low areas are less influent to determine the Natural Grasping Axis. Moreover the influence mark $\zeta_{i j}$ enables to balance out the relative weighting of unsatisfying couples of facing surfaces.

### 2.4 The iterative algorithm

When the direction of each axis is identified, it is essential to locate a point it passes through. This positioning phase consists in a normal projection of the centres of the influent facing surfaces on the plane perpendicular to this axis. It creates one or more classes of points according to the number of possible positions for the N.G.A. Then an algorithm calculates the separated gravity centres of each created classes to facilitate the localization of the grasping axis (Fig.6). The next decisive point consists in confining this infinite line into the geometrical envelope of the object to avoid positioning the hand into empty space (Fig.6).


Fig. 6. NGA positioning and confining

Moreover the optimal N.G.A. depends on the nature of the task and on the necessity to avoid obstacles. Subsequently, it is essential to retrieve a large panel of potential axes through an iterative process [5] before positioning them in the context of the 3D object. Thus, after the step of 3D acquisition, the algorithm identifies pairs of facing surfaces and generates the primary N.G.A.. Next, the influence of the surface couples $C_{a 1}$ contributing to the primary axis a1 is decreased by a reducing coefficient inserted in the quadratic criterion. Iterating this process leads to a set of possible grasping axis ranked in proportion to the surface areas they were generated by (Fig.7).


Fig. 7. Positioning and confining the panel of Grasping Axes

## 3 The 3D reconstruction of the object to grasp

Our grasp planner works with triangulated representations of the objects. Whereas shapes from silhouettes techniques [7] are easy to implement, they are inadequate to reconstruct the concavities of the objects. The methods derived from voxel coloring consist in carving a piece of voxelized virtual scene that contains the target object. This leads to a decomposition of the object into small elements of volume, called voxels. Of course the accuracy of the final 3D reconstructed volume depends on the number of involved voxels which can potentially extend the computational cost. Once these voxels are calculated, it is easy to obtain a satisfying triangulated representation of the object to determine the Natural Grasping Axis $[8]$.

### 3.1 Related Work

Dyer and Seitz [9] developed a method called Voxel Coloring, based on space carving, to reconstruct scenes with important color variation (Fig. 8). This method approaches the problem as a color reconstruction problem rather than a shape reconstruction to build photorealistic scenes.


Fig. 8. Example of a scene decomposed into colored Voxels

Voxel coloring requires the assignment of colors to points in a 3D volume. If a single point has the same color in all images from which it is visible, then it should be given this color. If the colors do not match, then the examined 3D volume is probably empty and the voxel should be removed. Therefore, if the voxel color value is similar in all the images, the voxel is consistent and will be kept. Otherwise, the voxel is rejected and removed from the global volume. The algorithm will stop when all the visible voxels are consistent with the images. Thus the consistency check [10] is essential to the voxel coloring algorithm. In order to take potential noises into account, the deviation dC is calculated for a given trivial color Red, Green or Blue:

$$
\begin{equation*}
d C=\sqrt{\frac{1}{K} \sum_{i=1}^{K} C_{i}^{2}-\left(\frac{1}{K} \sum_{i=1}^{K} C_{i}\right)^{2}} \tag{4}
\end{equation*}
$$

With K the number of pixels in the pixel collection of the footprint and Ci the value for one given color of the RGB color channels between 0 and 255. The main assumption is that all pixels of a voxel projection on a footprint must have the same color to be consistent. Therefore a voxel will be consistent if the sum $d R+d G+d B$ is inferior to a given threshold. However, the main problem is the difficulty to find an optimal threshold: areas with little texture are better reconstructed with a low threshold, while areas that are highly textured or contain sharp edges need high thresholds. A satisfying choice for the threshold enables to increase the robustness to varying illumination conditions between the different images.

### 3.2 The application to the N.G.A. extraction

In our experiment, the turntable is rotated to capture the video images of the real object in order to extract its contour. Therefore, a specific background (the pattern) was both used to calibrate the camera and to distinguish the object from the environment. Below we present the result (Fig.9) obtained by our algorithm with 1500000 voxels of 2 mm 3 for a real small teapot whose dimensions are sufficiently small (the radius of its body measures 11 centimeters) to be completely wrapped into our gripper. By rotating this object 360 degrees, 20 input images ( $640 \times 486$ resolution) were captured.


Fig. 9. The small teapot on the turntable

For the time being, the extrinsic calibration step requires to use a pattern grid and to select reference points in every image. A future crucial step will be to locate automatically such points in natural environments. Furthermore, between the calibration and reconstruction phases, there is a step of background removal which is far easier with an object placed on a showy pattern than when there is little variation in the background of the images taken. To reconstruct the following teapot (Fig.10) we used the Camera Calibration Toolbox for Matlab developed by Jean-Yves Bouguet [11] to retrieve the camera properties, position and orientation. Our reconstructions based on webcam images are surprisingly good, considering the image resolution thanks to the photorealistic quality engendered by Voxel Coloring. A major limitation is the fact that the visibility of each evaluated voxel has to be calculated which can


Fig. 10. The reconstructed teapot
be extremely time-consuming. Since the running time of the method depends on the number of voxels, a good estimation of the initial bounding box of the object must be made. Output of the camera calibration process gives a rough estimation of this volume. The efficiency is enhanced by using the next structure (Fig.11). The volume is refined after each step of this process. The


Fig. 11. How to increase the speed of 3D reconstruction
first step consists in discretizing the whole scene with big voxels to clearly identify the position of the object in the landscape. A bounding box is then calculated to decrease the computation time to a specific extent (up to 60\%) thanks to the decreasing resolution of voxels when approaching the target. Results show that for an object graspable by a human hand, voxels of 5 mm 3 are sufficient to extract a proper panel of N.G.A.. Actually even if this resolution is not sufficient to obtain a photorealistic representation of the object, it is adequate for our grasping topic. However, this method is limited to distinguish the target object from its obstacles in case of a cluttered environment. To solve this issue, an interesting suggestion lies in using, like Faugeras and Keriven [13], a variational formulation of the problem of 3D reconstruction such as a surface evolution given by Partial Differential Equations (P.D.E.).

## 4 Simulation Results

Considering that the voxel dimensions are small compared to the dimensions of the object, the set of involved voxels can be assimilated to a set of 3D points which enables to apply the Delaunay triangulation algorithm [12] in order to get the triangulated model of the object to grasp (Fig.12). Thanks to the set of calculated points, the plane can be split in a Voronoi diagram to obtain the final triangulated shape.


Fig. 12. The Delaunay Triangulation

One can notice that, with this triangulation method, the handle of the teapot is considered as a convex volume although this part of the global surface is composed of 300000 voxels. Keeping in mind that a Delaunay representation is unique, the same set of points will always produce the same triangulation irrespective of the order of which the data points are inserted. Consequently, the Delaunay algorithm is not completely satisfying in this case. Thus, it is essential to improve the quality of this triangulation with the purpose of solving issues caused by connectivity. After the phases of positioning, classification and confining, we finally obtain the panel of possible axes as illustrated in the following figure (Fig.13): The two first axes given by this


Fig. 13. The extraction of the N.G.A.
method are mainly generated by the revolution geometry of the body as well as the large area of the cover and of the base, respectively. The third axis is
surprising from the point of view of the task to be executed (pouring tea) but grasping the teapot by the spout remains nevertheless perfectly possible. The main difference with the results obtained with the virtual teapot [5] concerns the handle. Only one axis belongs to the handle with a worse ranking whereas there were two axes for the teapot modeled with 3Ds. As explained previously, this is due to the triangulated representation of the handle which is not precise enough even if the calculated axis remains acceptable for grasping. In fact, the convex hull engendered by this Delaunay triangulation does not give a proper representation of the connectivity of the handle.

## 5 Conclusion

In this paper, we have presented a grasping strategy for complex objects whose geometry is given by a real vision system. It results from an analysis based on the human grasping in the absence of obstacles and a survey of voxel coloring. Our grasp synthesis algorithm represents a novel approach for robotic grasping with any shape of object. In most cases, results fit with axes instinctively chosen by human beings. This method offers the advantage of a fast calculation. Actually a few number of voxels is sufficient to retrieve the virtual Natural Grasping Axes whose computation only depends on the minimization of a quadratic criterion. The future works mainly consist in implementing an algorithm providing a 3D geometrical perception of the target that improves as the hand approaches the object in a cluttered environment. Actually one of the major foreseen problem is that this methodology needs a complete representation of the geometry of the objects and obstacles. Yet, in real environments, obstacles can occlude a part of the target from this 3D vision system. Therefore the next step resides in getting proper Grasping Axes and contact points even with a partial representation of the object to grasp.

## Acknowledgment

The authors would like to thank Miss Kazuné Osato for her precious assistance to reconstruct a wide range of different 3D objects using Voxel Coloring with our practical turntable.

## References

1. Ch. Borst, M. Fischer, G. Hirzinger, May 1999, "A Fast and Robust Grasp Planner for Arbitrary 3D Objects", Proceedings of IEEE International Conference on Robotics and Automation, Detroit, Michigan, pp.1890-1896.
2. Ch. Borst, M. Fischer, and G. Hirzinger, October 2003, "Grasping the dice by dicing the grasp", Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3692-3697, Las Vegas, Nevada.
3. E. Lopez-Damian, D. Sidobre, and R. Alami, April 2005, "A grasp planner based on inertial properties", Proceedings of IEEE International Conference on Robotics and Automation, pp. 766-771, Barcelona.
4. J.M. Thomine, R. Tubiana, 1990, "La Main : anatomie fonctionnelle et examen clinique", Paris, éditions Masson, collection "Abrégés de médecine".
5. C. Michel, C. Rémond, V. Perdereau and M. Drouin, July 2004, "A Robotic Grasping Planner based on the Natural Grasping Axis", Proceedings of the International Conference on Intelligent Manipulation and Grasping, Genoa, Italy.
6. C. Michel, C. Rémond, V. Perdereau and M. Drouin, March 2004, "Robotic Fingers Positioning on Arbitrary 3D Objects", Proceedings of the International Symposium on Robotics ISR2004, Paris, France.
7. Y. Kuzu, 2002, "Photorealistic Object Reconstruction Using Color Image Matching", Proceedings of the ISPRS Commission Symposium, Close-Range Vision Techniques, Corfu, Greece, pp. 169-174.
8. C. Michel, C. Rémond, V. Perdereau and M. Drouin, September 2005, "Extraction of the Natural Grasping Axis from Arbitrary 3D envelopes provided by Voxel Coloring", Proceedings of the International Symposium of Robotics, Tokyo, Japan.
9. M. Seitz, R. Dyer, 1997, "Photorealistic scene reconstruction by voxel coloring", Proceedings of the IEEE International Conference of Computer Vision and Pattern Recognition, pp. 1067-1073.
10. A. Broadhurst, R. Cipolla, 2000, "A Statistical Consistency Check for the Space Carving Algorithm", Proceedings of the IEEE International Conference on British Machine Vision Conference, pp. 282-291.
11. Jean-Yves Bouget, 2004, Camera calibration toolbox for Matlab, http://www.vision.caltech.edu/bougetj/calib_doc.
12. C. B. Barber, D.P. Dobkin, and H.T. Huhdanpaa, December 1996, "The Quickhull Algorithm for Convex Hulls", ACM Transactions on Mathematical Software, Vol. 22, No. 4, pp. 469-483.
13. O. Faugeras and R. Keriven., 1997 , "Level set methods and the stereo problem", Proceedings of the First International Conference on Scale-Space Theory in Computer Vision, volume 1252 of Lecture Notes in Computer Science, pp. 272283.
