GRASPING WITH VISION DESCRIPTORS AND MOTOR PRIMITIVES

Oliver Kroemer
Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tuebingen, Germany

Renaud Detry, Justus Piater
University of Liege, Grande Traverse 10, 4000 Liege, Belgium

Jan Peters
Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tuebingen, Germany

Keywords: Dynamical Motor Primitives, Early Cognitive Vision Descriptors, Grasping.

Abstract: Grasping is one of the most important abilities needed for future service robots. Given the task of picking up an object from betweem clutter, traditional robotics approaches would determine a suitable grasping point and then use a movement planner to reach the goal. The planner would require precise and accurate information about the environment and long computation times, both of which may not always be available. Therefore, methods for executing grasps are required, which perform well with information gathered from only standard stereo vision, and make only a few necessary assumptions about the task environment. We propose techniques that reactively modify the robot's learned motor primitives based on information derived from Early Cognitive Vision descriptors. The proposed techniques employ non-parametric potential fields centered on the Early Cognitive Vision descriptors to allow for curving hand trajectories around objects, and finger motions that adapt to the object's local geometry. The methods were tested on a real robot and found to allow for easier imitation learning of human movements and give a considerable improvement to the robot's performance in grasping tasks.

1 INTRODUCTION

Consider the scenario wherein you want to have a humanoid robot grasp an object in a cluttered space. The first stage of most grasp planners determines a suitable grasp location on the object (Saxena et al., 2008; Arimoto, 2008; Bicchi and Kumar, 2000). Having selected a final location and orientation for the hand, the robot must then determine how to execute the grasp so as not to collide with the object or any of the surrounding objects.

The traditional solution for this scenario involves supplying the robot with a CAD model of the objects and a laser scanner or other means (ERFID, previous position, etc.) for obtaining their precise positions. These tools give the robot ample knowledge to apply a planning algorithm that determines a suitable path to the goal. This process relies on precise sensor information and can be very time consuming given a complex scene with numerous possible object collisions to test for at each step. In contrast, humans can perform successful grasps of objects in the periphery of their vision, where visual information is limited.

Taking inspiration from human movements, we propose a reactive method for robots grasping objects in cluttered environments using potential fields based on only a small amount of visual information. Specifically, we present methods for incorporating information derived from Early Cognitive Vision Descriptors (ECVD) (Pugeault, 2008) into the dynamical system motor primitives (DMP) (Schaal et al., 2003) framework. The Early Cognitive Vision system (see Appendix and Figure 2) was chosen since it makes only a few assumptions about the object being grasped, while the motor primitives (see Appendix) were chosen because they generalize well to new situations and can be learned through imitation (Ijspeert et al., 2002). The two frameworks are also compatible

Kroemer O., Detry R., Piater J. and Peters J. (2010).
GRASPING WITH VISION DESCRIPTORS AND MOTOR PRIMITIVES.
In *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics*, pages 47-54
DOI: 10.5220/0002938100470054

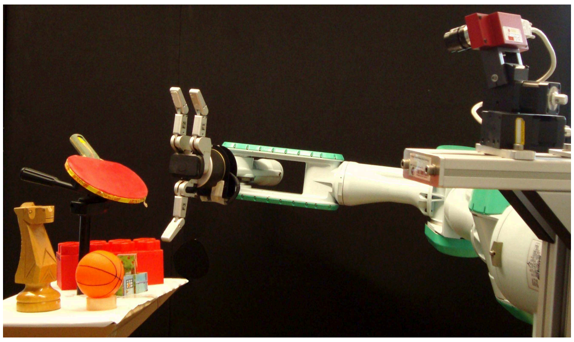

Figure 1: The robot used in our experiments and an example of a grasping task in a cluttered environment.

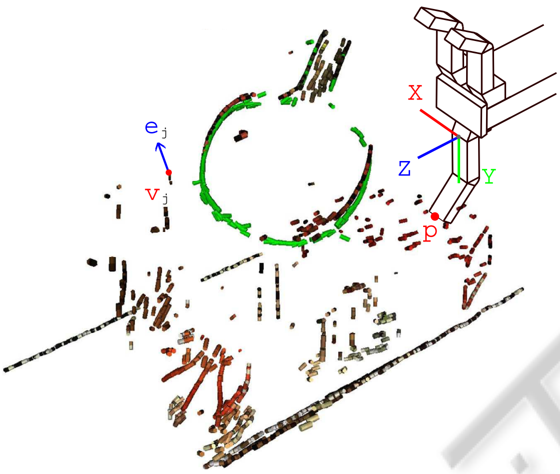

Figure 2: The green ECVD represent the object to be grasped, while the surrounding ECVDs in the scene are clutter. The coordinate frame of the third finger of the Barrett hand (the lower finger in the image) and variables used in section 2 are shown. The *x*-*y*-*z* coordinate system is located at the base of the finger, with *z* orthogonal to the palm, and *y* in the direction of the finger. The marked ECVD on the left signifies the j^{th} descriptor, with its position at $\mathrm{v}_j = (v_{jx}, v_{jy}, v_{jz})^T$, and edge direction $\mathrm{e}_j = (e_{jx}, e_{jy}, e_{jz})^T$ of unit length. The position of the finger tip is given by $\mathrm{p} = (p_x, p_y, p_z)^T$.

with each other and thus straightforward to combine.

The ECVDs were used to elegantly augment the DMPs for grasping tasks, resulting in the robot being able to avoid obstacles, curve its reaching trajectories around the object to grasp, and adapting the fingers to the local geometry of the object.

2 METHODS FOR REACTIVE GRASPING

The methods proposed in this section were inspired by human movements. Human grasping movements can be modeled as two linked components, transportation and finger posture, synchronized by a shared timer or canonical system (Chieffi and Gentilucci, 1993; Oztop and Kawato, 2009). Transportation refers to the actions of the arm in moving the hand, while the finger posture aspect relates to the preshaping and closing of the fingers (Jeannerod, 1997).

Humans perform the reaching/transportation component in a task-specific combination of retina and hand coordinates (Graziano, 2006), which allows for easier specification of object trajectories in a manipulation task than joint coordinates would and results in a reduction in dimensionality. These movements also have curved trajectories that are needed for avoiding obstacles and reaching around objects, which mainly occurs in a planar subspace (Wank et al., 2004).

Similar to the transportation component, the main purpose of the finger posture component is to preshape the hand by extending the fingers sufficiently for them to pass around the object upon approach, and then close on the object simultaneously for a good grasp. Over-extending the fingers is undesirable as it makes collisions with the environment more likely and is therefore usually restricted to situations that present large uncertainties about the object (Oztop et al., 2004; Chieffi and Gentilucci, 1993).

Curved reaching trajectories and preshaping of the hand were incorporated into the robot via a potential field, as described in Sections 2.1 and 2.2. Subsequently, a higher level controller is proposed in Section 2.3, which allows the grasping movements to be interpolated better to new target grasp locations.

2.1 DMP based Attractor Field

The first step towards specifying the grasping movements is to define an attractor field as a DMP that encodes the desired movements given no obstacles. The principal features that need to be defined for these DMPs are 1) the goal positions, and 2) the generic shape of the trajectories to reach the goal.

Determining the goal posture of the hand using the ECV descriptors has been previously investigated in. (Detry et al., 2009). In this work, possible grasp locations were hypothesized from the geometry and color features of the ECVDs, and used to create a kernel density estimate of suitable grasps, which is then refined by attempting grasps to test them.

However, this grasp synthesizer only gives the desired location and orientation of the hand, but leaves finger placement to a secondary finger controller, e.g., (Hsiao et al., 2009; Steffen et al., 2007). Using the ECVDs, the goal position of each finger is approximated by first estimating a contact plane for the object in the finger coordinate system shown in Figure 2. To make it a local approximation, the influence of the i^{th} ECVD is weighted by $w_i = \exp(-\sigma_x^{-2} v_{ix}^2 - \sigma_y^{-2} v_{iy}^2 - \sigma_z^{-2} v_{iz}^2)$, where σ_x, σ_y, and σ_z are length scale constants, and v_i is the position of the ECVD in the finger reference frame. The hand orientation was chosen such that the Z direction of the finger is parallel to the approximated contact plane, which reduces the problem to describing the plane as a line in the 2D X-Y space. The X-Y gradient of the plane is approximated by $\phi = (\sum_{i=1}^N w_i)^{-1} \sum_{i=1}^N w_i \arctan(e_{iy}/e_{ix})$, where N is the number of vision descriptors, and e_i is the direction of the i^{th} edge. The desired Y position of the fingertip is then given by

$$\tilde{p}_y = \frac{\sum_{i=1}^N (w_i v_{iy} - \tan(\phi) w_i v_{ix})}{\sum_{i=1}^N w_i},$$

which can be easily converted to a joint parameter using the inverse kinematics of the hand.

Having determined the goals of both transportation and finger-posture components, the next step is to define the trajectories used to reach these goals. Many of the beneficial traits of human movements, as described earlier, can be transferred to the robot through imitation learning. Learning by imitation involves a human demonstrating a motion and the robot then mimicking the movement. Details for imitation learning with DMPs can be found in (Ijspeert et al., 2002).

We can now combine the goals and imitation learned trajectories to specify the DMPs and thus the attractor fields.

2.2 ECVD based Detractor Fields

Having specified the rudimentary grasping movements, a detractor field is employed to refine the motions in order to include obstacle avoidance for the transportation and ensure that the finger tips do not collide with the object during the hand's approach.

The detractor field will be based on ECVDs, which can be envisioned as small line segments of an object's edges localized in 3D, as shown in Figure 2 for a scene as shown in Figure 1. The detractive potential fields for ECVDs are characterized by two main features; i.e., the detractive forces of multiple ECVDs describing a single line do not superimpose, and the field does not stop DMPs from reaching their ultimate goals. The system therefore uses a Nadaraya-Watson model (Bishop, 2006) of the form

$$u = -s(x) \frac{\sum_{i=1}^N r_i c_i}{\sum_{j=1}^N r_j},$$

to generate a suitable detractor field, where r_i is a weight assigned to the i^{th} ECVD, s is the strength of the overall field, x is the state of the DMPs' canonical system, and c_i is the detracting force for a single descriptor.

The weight of an ECVD for collision avoidance is given by $r_i = \exp(-(\mathrm{v}_i - \mathrm{p})^{\mathrm{T}} \mathrm{h}(\mathrm{v}_i - \mathrm{p}))$, where v_i is the position of the i^{th} ECVD in the local coordinate system, h is a vector of positive length scale hyper-parameters, and p is the finger tip position, as shown in Figure 2. The detractor therefore puts more importance on ECVDs in the vicinity of the finger.

The strength factor ensures that the detractor forces always tend to zero at the end of a movement and thus it can not obstruct the attractor from achieving its goal at the end. Therefore, the strength of the detractors is coupled to the canonical system of the DMP; i.e., $s(x) = (\sum_{j=1}^M \psi_j)^{-1} \sum_{i=1}^M \psi_i w_i x$, where x is the value of the canonical system, ψ are its basis functions, and w specify the varying strength of the field during the trajectory.

The transportation and finger-posture movements react differently to edges and thus employ different types of basis functions c_i for their respective potential fields. For the fingers, the individual potential fields are logistic sigmoid functions about the edge of each ECVD of the form $\rho(1 + \exp(d_i \sigma_c^{-2}))^{-1}$, where $d_i = \left\| (\mathrm{p} - \mathrm{v}_i) - \mathrm{e}_i (\mathrm{p} - \mathrm{v}_i)^{\mathrm{T}} \mathrm{e}_i \right\|$ is the distance from the finger to the edge, $\rho \geq 0$ is a scaling parameter, and $\sigma_c \geq 0$ is a length parameter. Differentiating the potential field results in a force term of

$$c_i = \rho \frac{\exp\left(d_i \sigma_c^{-2}\right)}{\left(1 + \exp\left(d_i \sigma_c^{-2}\right)\right)^2}.$$

As the logistic sigmoid is monotonically increasing, the detractor always forces the fingers open further to move their tips around the ECVDs and thus ensure that they always approach the object from the outside.

The hand uses instead the Gaussian basis functions of the form $\rho \exp(-0.5 \mathrm{d}_i^{\mathrm{T}} \mathrm{d}_i \sigma_d^{-2})$, where $\mathrm{d}_i = (\mathrm{q} - \mathrm{v}_i) - \mathrm{e}_i (\mathrm{q} - \mathrm{v}_i)^{\mathrm{T}} \mathrm{e}_i$ is the distance from the end effector position, q, to the edge, and $\rho \geq 0$ and $\sigma_d \geq 0$ are scale and length parameters respectively. Differentiating the potential with respect to d_i gives a force term in the Y direction of

$$c_i = \left[\rho \mathrm{d}_i \sigma_d^{-2} \exp(-0.5 \mathrm{d}_i^{\mathrm{T}} \mathrm{d}_i \sigma_d^{-2}) \right]_Y,$$

which can be interpreted as a radial force from the edge with an exponentially decaying magnitude.

The detractor fields, of both the grasping and reaching components, have now been defined, and can be superimposed into the DMP framework as

$$\ddot{y} = \left(\alpha_z(\beta_z \tau^{-2}(g-y) - \tau^{-1}\dot{y}) + a\tau^{-2} f(x)\right) - \tau^{-2} u,$$

which then represents the entire ECVD and MP based potential field.

2.3 High Level DMP Controller for Grasping

Having defined the potential field for a single grasping motion, we interpolate the movements to new target grasps. Having a motion representation that can be interpolated to new targets is crucial for imitation learning. Given such a representation, the number of example trajectories required from the demonstrator can be greatly increased, making learning easier. While DMPs can interpolate to arbitrary goal positions, they have two drawbacks for grasping tasks; i.e., 1) the approach direction to the grasp can not be arbitrarily defined, and 2) the amplitude of the trajectory is unneccessarily sensitive to changes in the start position y_0 and the goal positon g if $y_0 \approx g$ during training, which can cause the robot to reach the limits of its workspace.

These difficulties can be overcome by including a supervisory controller that modifies the hyperparameters of the DMPs appropriately. The supervisor can maintain the correct approach direction by using a task-specific coordinate system. Due to the translation invariance of DMPs, only a rotation, $\mathrm{R} \in \mathbb{SO}(3)$, between the two coordinate systems needs to be determined. The majority of the motions will lie in a plane defined by the start and goal locations, and the final approach direction.

The first new in-plane axis x_p is set to be along the approach direction of the grasp; i.e., $\mathrm{x}_p = -\mathrm{a}$ as shown in Figure 3. As a result, the approach direction is easily defined and only requires that the Y_p and Z_p primitives reach their goal before the X_p primitive. The second axis, y_p, must be orthogonal to x_p and also in the plane, as shown in Figure 3. It is set to $\mathrm{y}_p = b^{-1}((\mathrm{g}-\mathrm{s}) - \mathrm{x}_p(\mathrm{g}-\mathrm{s})^T\mathrm{x}_p)$, where b^{-1} is a normalization term, and s and g are the motion's 3D start and goal positions respectively. The third vector, z_p, is orthogonal to the plane, and is derived by completing the right-handed coordinate system, i.e., $\mathrm{z}_p = \mathrm{x}_p \times \mathrm{y}_p$. The DMPs can now be specified by the supervisor in the X_p-Y_p-Z_p coordinate system, and mapped to the X_w-Y_w-Z_w world reference frame by multiplying by $\mathrm{R}^T = [\mathrm{x}_p, \mathrm{y}_p, \mathrm{z}_p]^T$.

The second problem relates to the scaling of motions with ranges greater than $y_0 - g$, which both com-

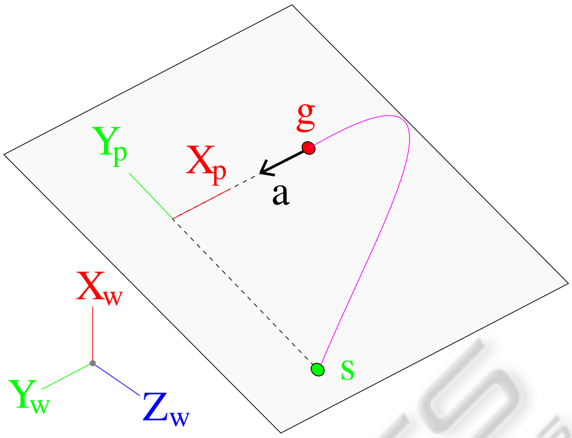

Figure 3: The above diagram shows the the change in coordinate systems for the transportation DMPs. The axes X_w-Y_w-Z_w are the world coordinate system, while X_p-Y_p-Z_p is the planar right handed coordinate system in which the DMP is specified. The trajectory of the DMP is shown by the **pink** line, starting at the **green point**, and ending at the **red point**. Note that X_p is parallel to the approach direction of the hand, which is shown by the **black arrow** a. The planar axis Y_p is perpendicular to X_p, and pointing from the motor primitive's starting location s towards the goal g.

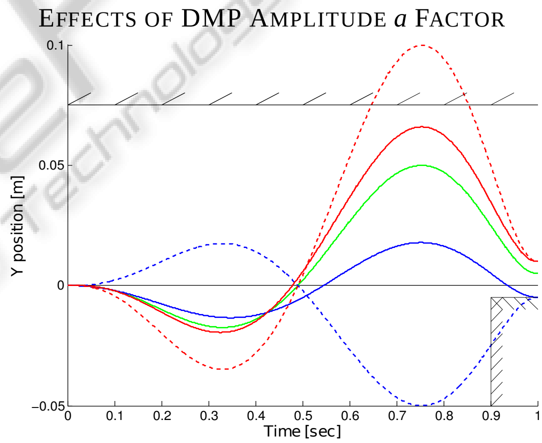

Figure 4: This is a demonstration of the effects of augmenting the amplitude variable a of DMPs. The black lines represent boundaries. The **green** plot shows the trained trajectory of the DMP going to 0.05, and is the same for all amplitude values. Now consider the scenario wherein the goal is placed at 0.1, but the workspace is limited to ± 0.75 (top boundary). The **dashed red** line is the standard generalization to a larger goal, while the **red** plot uses the new amplitude. Notice how the new amplitude restricts the range of the trajectory to the workspace. In a different scenario, we move the goal to -0.05, but require the goal to be reached from above (lower right boundary), e.g., a finger placed on a surface. The **dashed blue** line is the standard generalization to a negative goal, and the **blue** trajectory uses the new amplitude. Note that the trajectory is not flip in the case of the new amplitude and thus stays within the restricted region. Both of the new trajectories were generated with $\eta = 0.25$, and maintain shapes close to that of the training trajectory.

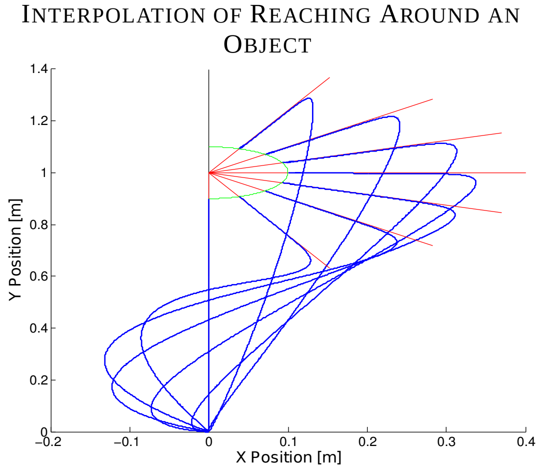

Figure 5: The plot shows workspace trajectories, wherein the x and y values are governed by two DMPs sharing a canonical system. The **red** lines indicate the desired approach direction while the **green** semicircle indicates the goal positions along them. The **blue** lines show the trajectories for the different goals. They make use of the higher level controller of Subsection 2.3, with $\eta = 0.25$. The approach direction DMP was trained on an amplitude of one.

ponents require to move around the outside of objects. In the standard form $a = g - y_0$ (Ijspeert et al., 2003), which can lead to motions that easily exceed the robot's workspace if $g \approx y_0$ during the training, but not during the motion reproduction. The supervisor can control these trajectories by scaling the shaping force (see Appendix), and thus we propose the amplitude term

$$a = \|\eta(g - y_0) + (1 - \eta)(g_T - y_{0T})\|,$$

where g_T and y_{0T} are the goal and start positions of the training data respectively, and $\eta \in [0, 1]$ is a weighting hyperparameter. The resulting trajectory amplitude is in the convex hull of the training amplitude and the standard interpolation value ($a = g - y_0$) (Ijspeert et al., 2003) and thus only affects how conservative the generalization to new points is, as can be seen in Figure 4. By taking the absolute value of the amplitude, the approach direction is not reversed, giving a result similar to the use of a constant amplitude proposed by Park et al. (Park et al., 2008), which corresponds to the special case of $\eta = 0$.

Example interpolations of a transportation trajectory can be seen in Figure 5.

3 GRASPING EXPERIMENTS

The methods described in Section 2 were implemented and evaluated on a real robot platform. The robot consists of a Videre stereo camera mounted on a pan-tilt unit, a Barrett hand, and a Mitsubishi PA10 arm. The robot was given the task of grasping an object amongst clutter using only an ECVD model of the object. The results of these trials were then compared to trials of the same grasps using other standard robotics methods for comparison. We hypothesize that our method will result in significantly more successful grasps than the other methods.

3.1 Grasping Experiment Procedure

Before the robot can perform a grasping task, its motions must be initialized. Determining the finger goal state and specifying the detractor fields introduces several new hyperparameters that have simple geometrical interpretations. For instance, $\mathrm{h} = 2[w\ l\ l]^{\mathrm{T}}$, where w and l are the width and length of the finger respectively. To reflect the human tendency towards more precise movements during the last 30% of a motion (Jeannerod, 2009), the strength function, $s(x)$, was set to give the highest strengths during the first 70% of the motion for the transportation, and the last 30% for the finger posture.

A VICON™ motion tracking system was used to record the movements of a human test subject during a grasping task, which used a different object to the one used by the robot. As the reaching trajectories are encoded in task space rather than joint space, the correspondence problem was not an issue for the imitation learning. Similarly, the DMPs of the fingers are homogeneous, which circumvents the correspondence problem of mapping the five human fingers onto the three fingers of the robot. The imitation learning was performed using locally weighted regression in the the X_p-Y_p-Z_p coordinate system, as proposed by Ijspeert et al. (Ijspeert et al., 2002).

Having defined the basic motions, the robot was then given the task of grasping an object without hitting surrounding obstacles (see Figure 1). Each trial begins with an estimate of the pose of the object relative to the robot (Detry et al., 2008) and sets its grasp location accordingly. The model's ECVD are then projected into the scene, and the robot attempts to perform the grasp and lift the object 15cm so that it is clear of the stand. The trial is a success if the robot can detect the object in its hand at this point. If the hand collides with an obstacle or knocks the object down, the trial is marked as a failure. Grasps were varied to include different approach directions and locations around the object. The experiment consisted of 45 trials.

Two alternative approaches were compared with our proposed method. The first represents a standard robotics approach of specifying a trajectory by

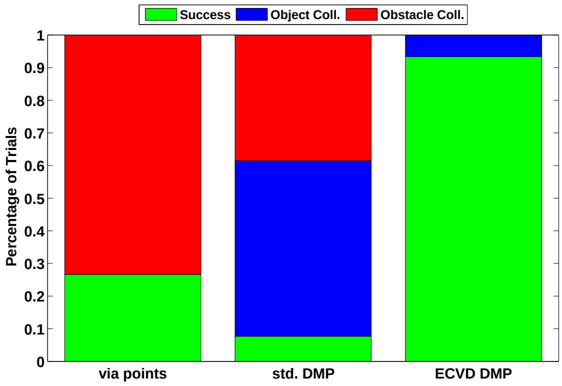

Figure 6: The occurrences of successes and collision types for the different methods are shown. The first column presents the results for the traditional robotics method of specifying trajectories by via points. The second column corresponds to using standard DMPs, while the final column incorporates the ECVD based potential field and supervisory DMP controller. The occurrences are given as the percentage of trials. Trials that collided multiple times, are classified by their first collision.

straight lines between via points and uses fully extended fingers with no preshaping of the hand. The other approach is to use standard DMPs learned from the same human demonstrated movements as our proposed methods, but without the proposed detractor field and supervisory controller. The same grasp locations were proposed to the different methods, and obstacles were placed in similar positions for the the different trials to allow for a fair comparison between the methods.

3.2 Experimental Results

From the three tested methods, the proposed method acquired the highest success rate, as can be seen in Figure 6. The task was not trivial, and all of the methods encountered both successes and problems during the trials.

The standard DMP method encountered the most problems (success rate of only 7%) a majority of which were caused by collisions with the object. This high failure rate can be attributed to the method not specifically incorporating a desired approach direction. In successful trials, the approach direction was close to that of the initial imitation learning. Therefore the proposed DMP supervisor improved the generalization of the movement to new target grasps, and the system would benefit from it even in uncluttered environments. Similarly, the open-loop preshaping of the hand helped avoid obstacles, but occasionally prevented the hand from being sufficiently open to accept the object. The proposed detractor field successfully overcame this problem for the ECVD DMPs.

The via points method encountered no collisions with the object, and would have worked well in an uncluttered environment. The method still encountered collisions with the obstacles for 73% of the trials, but this is more reflective of the difficulty of the task rather than the via point method. The method can therefore be considered as a good approach if it were combined with a suitable path planning method for obstacle avoidance. However, the path planner would need additional information and assumptions about the scene and possibly even extra hardware to acquire it.

The proposed method had a success rate of 93%, with no occurrences of collisions with obstacles. The trials that did fail were the result of the object falling down while the fingers were closing and thus do not indicate problems with the approach used to reaching the grasp location. The method does have certain restrictions though. The magnitude of the detractor fields needs to be calibrated based on the density of ECVDs for common objects, but some obstacles encountered may present lower densities. As the current set of ECVD relies on object edges, smooth objects can lead to noisy or very sparse descriptors, and therefore not create a suitable basis for obstacle avoidance. As the number of descriptor types increases (e.g., corner and plane descriptors), this will become less of a problem. Occluded obstacles will also need to rely on additional information (e.g., force feedback) to be avoided, although this is a source of error for all vision based planners.

Given a few restrictions, the results still show that our hypothesis was correct and the proposed methods represent a suitable basis for avoiding obstacles without relying on a complicated path planner and using only a small amount of vision information compared to standard robot systems.

4 CONCLUSIONS

The proposed methods augment dynamical system motor primitives to incorporate Early Cognitive Vision descriptors by using a potential field. These methods represent important tools that a robot needs to reactively execute grasps of an object in a cluttered environment without relying on a complex planner. The techniques allow for preshaping the fingers to match the shape and size of the object and curving the trajectory of the hand around objects(Wank et al., 2004). These modifications were tested on a real robot, and it was discovered that the methods were not only successful at performing the task, but also allowed for easier imitation learning, better interpolation of the learned trajectories, and significantly better chances of a success of a grasp in cluttered environments than standard motor primitives. Although the experiments were performed within a grasping task scenario, the proposed methods can be beneficial for other manipulation tasks, such as pressing buttons and pushing objects.

REFERENCES

Arimoto, S. (2008). *Control Theory of Multi-fingered Hands*. Springer London.

Bard, C., Troccaz, J., and Vercelli, G. (1991). Shape analysis and hand preshaping for grasping. In *Proceedings of IROS'91*.

Bicchi, A. and Kumar, V. (2000). Robotic grasping and contact: a review. In *ICRA 2000 proceedings*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Chieffi, S. and Gentilucci, M. (1993). Coordination between the transport and the grasp components during prehension movements.

Detry, R., Kroemer, O., Popovic, M., Touati, Y., Baseski, E., Krueger, N., Peters, J., and Piater, J. (2009). Object-specific grasp affordance densities. In *ICDL*.

Detry, R., Pugeault, N., and Piater, J. (2008). Probabilistic pose recovery using learned hierarchical object models. In *International Cognitive Vision Workshop*.

Graziano, M. S. (2006). Progress in understanding spatial coordinate systems in the primate brain. *Neuron*.

Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hsiao, K., Nangeroni, P., Huber, M., Saxena, A., and Ng, A. (2009). Reactive grasping using optical proximity sensors. In *ICRA 2009 Proceedings*.

Iberall, T. (1987). Grasp planning for human prehension. In *Proceedings of ICAI'87*.

Ijspeert, A. J., Nakanishi, J., , and Schaal, S. (2002). Movement imitation with nonlinear dynamical systems in humanoid robots. In *ICRA*.

Ijspeert, A. J., Nakanishi, J., , and Schaal, S. (2003). Learning attractor landscapes for learning motor primitives. In *NIPS*.

Jeannerod, M. (1997). *Perspectives of Motor Behaviour and Its Neural Basis*, chapter Grasping Objects: The Hand as a Pattern Recognition Device.

Jeannerod, M. (2009). *Sensorimotor Control of Grasping: Physiology and Pathophysiology*, chapter The study of hand movements during grasping. A historical perspective. Cambridge University Press.

Krueger, N., Lappe, M., and Woergoetter, F. (2004). Biologically motivated multimodal processing of visual primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*.

Oztop, E., Bradley, N. S., and Arbib, M. A. (2004). Infant grasp learning: a computational model.

Oztop, E. and Kawato, M. (2009). *Sensorimotor Control of Grasping: Physiology and Pathophysiology*, chapter Models for the control of grasping. Cambridge University Press.

Park, D.-H., Hoffmann, H., Pastor, P., and Schaal, S. (2008). Movement reproduction and obstacle avoidance with dynamic movement primitives and potential fields. In *IEEE International Conference on Humanoid Robots(HUMANOIDS)*.

Pugeault, N. (2008). *Early Cognitive Vision: Feedback Mechanisms for the Disambiguation of Early Visual Representation*. Vdm Verlag Dr. Mueller.

Saxena, A., Dreimeyer, J., Kearns, J., Osondu, C., and Ng, A. (2008). *Experimental Robotics*, chapter Learning to Grasp Novel Objects using Vision. Springer Berlin.

Schaal, S., Peters, J., Nakanishi, J., and Ijspeert, A. (2003). Learning movement primitives. In *Proceedings of ISRR'03*.

Steffen, J., Haschke, R., and Ritter, H. (2007). Experience-based and tactile-driven dynamic grasp control. In *IRS proceedings*.

Wank, V., Fischer, A., Bos, K., Boesnach, I., Moldenhauer, J., and Beth, T. (2004). Similarities and varieties in human motiontrajectories of predefined grasping and disposing movements. *International Journal of Humanoid Robotics*.

APPENDIX

Dynamical Systems Motor Primitives

The dynamical systems motor primitives (DMPs) proposed by Ijspeert et al. (Ijspeert et al., 2003) were inspired by the simple, but highly adaptive, motions that animals employ, and combine to obtain more complex motions. The primitives are implemented as a passive dynamical system with an external force, and represented as

$$\ddot{y} = \alpha_z(\beta_z\tau^{-2}(g-y) - \tau^{-1}\dot{y}) + a\tau^{-2}f(x), \quad (1)$$

where α_z and β_z are constants, τ controls the duration of the primitive, a is an amplitude, $f(x)$ is a nonlinear function, and g is the goal for the state variable y.

By selecting α_z and β_z appropriately, and setting $a = 0$, the system reduces to $\ddot{y} = \alpha_z(\beta_z\tau^2(g-y) - \tau\dot{y})$ and becomes a critically damped global attractor. It can be visualized as a spring and damper system that ensures state y will always end at the goal value g.

The function $f(x)$ is a shaping function based on the state, $x \in [0,1]$, of the canonical system that synchronizes the DMPs $\dot{x} = -\alpha_x\tau x$, where α_x is a time constant. The function takes the form

$$f(x) = \frac{\sum_{j=1}^{M} \psi_j(x) w_j x}{\sum_{i=1}^{M} \psi_i(x)},$$

where M is the number of basis functions, $\psi(x)$ are Gaussian basis functions, and w are weights acquired through locally weighted regression (Ijspeert et al., 2003). This function has the effect of introducing a non-linearity that can affect the spring-damper system to output any arbitrary trajectory specified by the user. Due to the dependence of $f(x)$ on x, the shaping term decays to zero with x, so that the spring and damper beneficial properties of the attractor are maintained.

The resulting primitives can encode arbitrary trajectories, and still ensure that the goal state is always achieved. The trajectories can also be scaled in time and space by setting the τ and g variables appropriately and thus generalize to a range of situations.

Early Cognitive Vision System

The entire prehensile process effectively occurs before the hand has even touched the object and thus the vision system plays a very important role (Bard et al., 1991; Iberall, 1987). Our system uses the Early Cognitive Vision methods of Pugeault et al. (Pugeault, 2008; Hartley and Zisserman, 2000), which makes a minimal number of assumptions about the object, and has been successfully implemented to determine good grasp locations (Detry et al., 2009). A principal idea of this vision system is to store additional low level information and perform perceptual grouping on it to later aid the higher level stereo matching and 3D constructions.

The methods extract local features of a scene, which it localizes and orientates in space (Krueger et al., 2004). Each descriptor is a symbolic representation for an edge in 3D. The resulting features are called early cognitive vision descriptors (ECVD) (Pugeault, 2008), and can be used in generating models of objects for pose estimation (Detry et al., 2008), and for symbolically describing 3D scenes. By using a large amount of small ECVDs, any arbitrary object can be represented.

When performing a grasping task, the robot uses a hierarchical Markov model of the object's ECVD geometry (Detry et al., 2008) to determine its pose, which can then be used to superimpose the ECVDs of the model back into the scene. The grasping techniques can therefore use geometric information of a partially occluded object.