Technical Writing

EST 304

Research Paper:

AI in Character Animation

Spring 2019

Professor:

Krista Thyberg

Author:

Sudhin Domala #110475495

Table of Contents

I. Abstract

The central theme of my research paper is the incorporation of Artificial Intelligence in character animation via machine learning. This discipline of animation is geared toward the production of special stylistic effects for films and interactive films like video games. Composed of software engineers, fine artists, and animators, computer graphics is a popular medium of many scientific and artistic disciplines. However, there is an existence of an artistic bottleneck where there are compromises to the efficiency and scalability of machine-generated models. (Alfonsi 2004, 4) This problem for the constant need for human-intervention for computer graphics can be solved with this research's computational strides in algorithm reusability. This need for synthesizing and generalizing data extends to various sectors of computer graphics from modeling, rendering and animating. Thereby, this paper's core objective is to illustrate the use of new ML methods (ex: reinforcement learning) to distinguish the pace/accuracy of character model automation compared to the required human workload for a finishing "film" product. Animators can have their characters based on different parameters to project a scene. They can be done in a speech-based, physics-based, or behavioral based manner. Due to their wider impact on the industry, Machine learning implications on these manners will first be explored. The second sphere of character creation that this paper will later explore will be around Rendering for project-specific automation. The three sectors of computer graphics will have their own introductory texts and constraints that machine learning can aim to leverage.

Having such artistic endeavors based on Computer-generated images/models are already implemented in practice and successful behind the creators of "Spiderman-In to the Spiderverse." Their animation tech consists of the same technological concepts to compute the artistic expression that the movie was awarded for as 2019's "Best Animated Picture". The reinvention of the animation pipeline

leads to these creators having meticulous amounts of time to invent the tools and implement them into practice via the pace of one second of screentime to a week's time of work. Sony also seeks to patent its Ink-line software application as cited "stylized abstractions of reality" and a machine-learning component of their animation process. Citing this instance of revamping the animation process indicates the extent that this research proposal can go in real-life applications. For instance, the quality of rendering and animation output has numerous factors involved in the hardware (image resolution, pixel depth, shading algorithms, and anti-aliasing methods) and software (ray-tracing, shadow casting, texture mapping, reflection mapping, shading algorithms). Current PC machines are still poorly supported to make such processes onto large 3D models or animation systems due to the long computational time alone. (Vasilopoulos 1990) Such systems incentivize artists to cut corners like reducing the resolutions of individual frames (thus can have unpleasant motion stills).

Augmenting from the concepts used in current industry's animation process, this research seeks to have an ideal learning-based animation system that allows the user (be whatever technical or artistic role they play in the workplace) "to supply a set of reference motions for style and then generate goal-directed and realistic character behavior from those reference motions." (Peng and Van De Panne 2018, 2) For such software to be realistic in a professional setting, some may say graphics techniques don't have to be (and sometimes can't be) solely determined with machine learning. Even so, Reinforcement learning does indicate a promising approach for such motion synthesis via an agent that learn to perform various skills that ultimately reduce human insight. This research will continue to undermine the existence of the problems in Reinforcement learning such as extraneous artifacts, unrealistic motions in the provided motion capture or keyframed animations. Such insight in a technology's capabilities will be elaborated to regression and data extrapolation concepts yet to be researched.

II. Technologies and Innovative Procedures

A. Machine Learning in Animation

1. Speech Recognition

Many current-gen graphics and Hollywood CGI equipment have computer vision where they take a grid-like input to animate a face in highly realistic proportions on screen. Facial Realism is critical to detect underlying expressions and emotional depth of the character in the scene. Give such a range, human viewers can easily detect incorrect sound-to-model correlations in the rendering schemes. Thus, voice puppetry is a technique that modern animators utilized "to track the facial features of a human demonstrator with a predefined script."(Dinerstein and Cline 2006, 31) The motion data is comprised of models that take in different probabilities of transitions amongst the facial expressions and rates. This is where a lot of artistic touches or corrections could to pick and choose the most optimal form for presentation. Machine Learning in this sector can automate the synthesis to yield faster, and self-reliant process of proper frame selection. It's worth to mention that this will pave an opportunity for interactive application of rapid online speech animation (ex: virtual reality or augmented reality). Machine Learning methods in data extrapolation are within the constraints of modern-tech limitations compared to the possibilities that ICA (independent component analysis) and regression can partake in. This conceptual application is in dire need of prior development compared to its counterparts in physics engines and behavioral systems that are already in place.

IEEE Intelligent Systems made a great launching pad for the cohesion of these concepts in a example-based learning method based on a video sequence of facial behavior like lip motion to define MPEG-4 face animation parameters. The method will consist of three key modules: face tracking, pattern learning, and face animation. (Gao and Jiang 2003) Overall this is an iterative process to reduce the tracking process of major landmarks in each image of a video reel. The Making of Miles Morales, the main character in Into the Spiderverse, was in a similar manner of taking a basic automated expression and adding extra nuances from more wrinkles, and other stylistic misplacements. (Figure 1) The automation is based on a visual feature matrix that is fed to basic mapping training. The parameters of the camera (in this case the perspective of the rendering software) and computer vision methods (ex: a Kalman filter) will estimate the normalization of the vivid expectations fluid facial motion. All in all, the system can ideally synthesize face animation from any time frame of the given video (such a specific set of surprise faces). (Gao and Jiang 2003) The implications can automate frame selection for a character to express a series of emotions.

Figure 1:

Major constraints will on the application level of capturing and modeling certain mouth structures to other diverse problems. Problems like timing hands or other bodily movements in the correlation of a specific facial expression will be critical to evaluate whether this process can be implemented on a large scale. Motion capture for such small gestures is difficult due to their own increasing degrees of freedom. This gives more room of noticeable areas of incorrect motion and again highlight the need to parameterize the different styles of speech motion.

2. Physically-based Animation

This style of rendering comes from simulating the physics in an animation system for specific graphical objectives. For instance, if an animator needs fluid flow or proper correlated-clothing movement then there are many challenges he/she faces. Lack of animator control, reduced computation time, and design difficulties cover the range of such challenges in this sphere. (Dinerstein and Cline 2006, 32) Machine Learning (esp neural networks) will be trained to pre-compute the necessary parameters for physics system-based simulations. Building on Neuro Animator (a neural net that is trained on system behavior to provide state transitions) or making general approximating stochastic systems will tabulate the path for system excitation. In essence, machine learning can vastly build the system specifications for proper animation prior to the artist's improvement for fluid flow of deformable (malleable) objects. Objects like human skeletons, swaying hair or erupting clouds can be analyzed to what approximate level of pruning is needed for the number of collision tests to be visually sufficient.

As said before, physics simulations are often difficult to design, code and execute on the desired timeline for the scene that the animator has in mind. So existing techniques can be an additional layer of inner works for motion capture like a model suit for a humanoid model to mimic the physical actions set by the animator. For instance, Figure 2 demonstrates a humanoid model imitating a physical system of a cartwheel while the robot performs a spin work with the same base system but different parameters. Working on physics system-based simulations can be worthwhile with its traversability to "multiple characters (human, Atlas robot, bipedal dinosaur, dragon) and a large variety of skills, including locomotion, acrobatics, and martial arts." (Peng and Van De Panne 2018, 1).

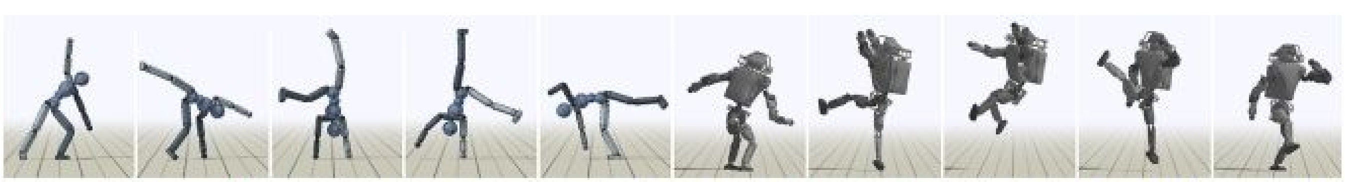


Fig. 1. Highly dynamic skills learned by imitating reference motion capture clips using our method, executed by physically simulated characters. **Left:** Humanoid character performing a cartwheel. **Right:** Simulated Atlas robot performing a spinkick.

The input layer of the neural network will include other properties of these characters like Total mass, Height, degrees of freedom (moving joints), state features, and action parameters.

The biggest constraints will be around where do the task objectives and termination policies are defined in the physics simulations. There are too many specifications where the base engineering knowledge required may not exist with the current animator's skill sets. While the majority of this knowledge lies around motor controllers, there is a need for techniques that automatically learns the circumstances of which controllers are useful for a character. Techniques like these will be beneficial due to the difficulty and time-consuming process that the animator would have to go through. Research attention in these learning-base approximation techniques needs to be leveraged to undermine these constraints. (Dinerstein and Cline 2006, 33) Finally, utilizing real-world examples through computer vision is an inspiring concept to model fluid flow from other videos. This pushes the limitations of machine learning like the input for this software can be more concise instead of an ambiguous amount of physical attributes.

3. Behavioral Animation

The most multi-disciplinary approach to animation is around behavioral procedures. Automation is this process will be around the concept of making self-animating virtual characters via making them as autonomous agents. These agents initiate its own decisions to define a character's thought process in a specific scene. This conceptual manner of animation overlaps many other fields like robotics and other AI systems compared to the use of physics or speech primarily around film and thus video games. The primary motives for this approach come for more realistic behavior and reducing programmer workload. (Dinerstein and Cline 2006, 33). Artificial Intelligence like Q-learning or an evolved neural network in this model format has currently very limited success with optimization problems. Ideally, the input for this software is in state-action pairs where certain behavioral patterns are compared to a certain situation.

Thus neural networks are around function-approximation around the character's decision-making process. When the scenic features are small and not iterative, this has worked well even in online platforms for real-time mimicry due to the offline basis of machine learning.

For instance, Figure 3 has a first real-time simulation of flying a satellite through asteroid fields by a real spaceship pilot. The program takes this experience as input in a timeline basis to perform its own traversal through a different set of asteroid fields. As its primary objective is to maneuver with no collisions, this is overall a master-slave relationship between the user and character (satellite model) respectively. The user's primary input provides feedback on which desirable behavior is exhibited and in essence augment the software's "memory. Remembering given instructions and information, there is room for questions or a interactive interface for AI improvement. All in all, interactive films from 4-D feature rides in amusement parks to virtual reality can greatly benefit from these intelligent systems. Training programs and videogames can be more realistic where character animation can recognize other forms of user input to make the experience more immersive.

Figure 3:

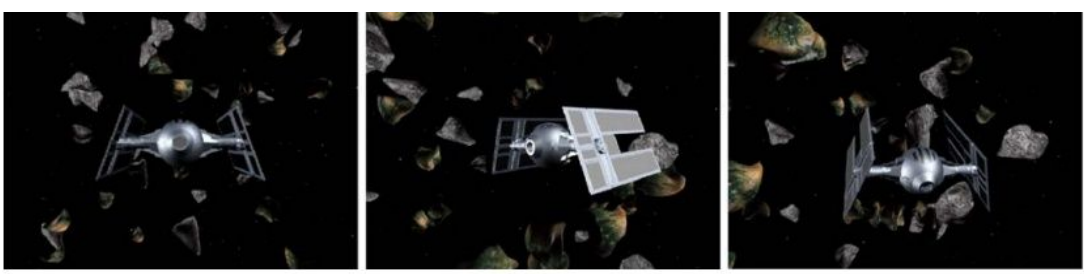


Fig. 12. Example of behavioral animation. The space ship pilot learns offline how to navigate asteroid fields. It then autonomously maneuvers through an unknown asteroid field with no collisions. From Dinerstein et al. [28] (courtesy of Wiley)

Despite many of the computational constraints, behavioral animation can be very useful/immersive in entertainment environments. Thus, there are various open avenues for future work to provide models with the power of observation of real-world behavior, allow smart characters to prioritize its interactions to its given environment, and scalable pre-computed behavior models. (Dinerstein and Cline 2006, 34). This type of animation is the most applicable use with these types of models but the actual construction is a daunting task. Research-focused endeavors will be critical after the accumulation of other forms of animation optimization has taken place. While being the most flexible of the three topics, actual commercial implications will stem on the product cycle that the company wants to undertake. User feedback can inhibit creative or stylistic preferences that animators would want to go for in these models. This software can show many bugs or mishaps that are not in the line of the lead animator (manager of the animation team) and even be difficult to find/resolve compared to a well-defined physics and speech automation systems.

B. Machine Learning in Rendering

1. Screen Space Ambient Occlusion

Compared to animation, rendering has very little breakthroughs incorporating machine learning to optimize its workflows. However, exploring this field can prove very beneficial in bridging the AI concepts with practical software learning methods. Lighting and Rendering play a huge role in inhibiting the atmosphere of the animated film. It spans from textures, setting the mood, have practical effects (fog,rain,sunlight) to insinuate the scene's background. Without rendering, animation of these character models will look slate or inaccurate compared to the nuances of the background. Animating fur, skin, metal shine are some of many computational obstacles animators have to exhibit physical attributes in their character models.Loading the textures faster can streamline the process of achieving the right portrayal that the animator is going for.

For this section, Ambient Occlusion will be touched on when lighting the scene to what animator preference. Due to Ambient Occlusion being expensive to calculate, many animators are discouraged to implement its techniques in the workplace. Machine learning can stiffen this dilemma with first databases of camera depths, normals, and ground truth ambient occlusion acting as inputs for the neural network. One can then train the network to learn a mapping of the database attributes around the desired pixel on the screen. We can later optimize the network to be a way better shader compared to the current ones in the market (due to better preference, and less user parameters). (Holden and Komura 2016) This incorporation of machine learning enables more areas of application due to its new-found data independence. A greater level of parallelism, better IO performance, less noise are all benefits that come with the cost of more required computation. All in all, Figure 4 does so the sufficient improvement in rendering with machine learning (right-side models) then without (left-side models).This threshold can be a reoccurring constraint to the today’s animation studios and thus another technological bottleneck. Higer details compared to the rendered models in Figure 4 will be a new realm of training for the data at hand. For instances, more colors, light specifications, or extra conflicting artifacts are other animation features to consider. All in all, constraints come from the need of more control over performance in these AI methodologies.

Figure 4:

Figure 1: *Comparison showing Neural Network Ambient Occlusion enabled and disabled inside a game engine.*

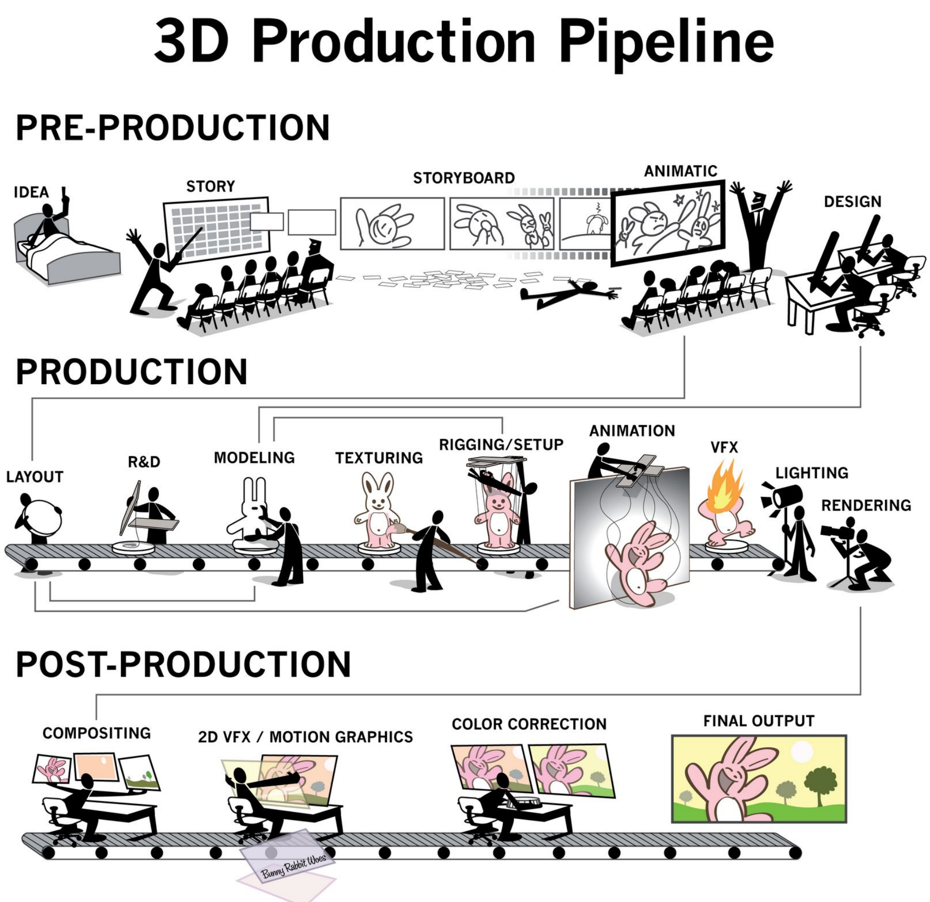
Scene complexity is a recurring factor for machine learning in animation due to its ambiguity but need of simple implementation for industrial use. Compared to other animation techniques and modeling, this factors more to be the proper catalyst for better automation in an animator’s rendering process. The biggest strength for machine learning in this field is that it's easier to incorporate older techniques like Screen Space Ambient Occlusion (SSAO) then building automation systems from scratch like the others.

2. Overall Rendering Technology

As mentioned before about scene complexity, rendering quality is widely depending on a number of factors in the animator's workstation. These include hardware and software components brought up before and their respective computational processes require arithmetic-intensive operations that are not runnable in average animator's three-dimensional models. Due to this constraint, there is an "tendency to minimize the number of objects in the model to be animated, and to reduce the resolution used to produce the individual frames well below what would normally be used for a single rendering." (Teicholz 1990). In the current climate, optimizing the resolution is a skill within the artistic bracket of the animator rather the one's technical ability due to the project-specific subject matter. Lighting and distance of a virtual camera from the target character is too played out in the intending of the animation director then some technical preference. This style threshold is the biggest constraint for machine learning to be part of the usual animation procedure.

Resource management for all this computation is a big concern to be mentioned for these technologies as a whole. Especially in rendering, hardware and software technologies are costly and have to come in high volumes to anticipate this automation. The human workload to for making a model along with the render specifications (character by character and scene by scene) is generally still in a reasonable time compared to this AI software development. This conversion process comes with proper management and oversight from design architects (engineers) to be a controllable and quality-assured process of animation. As seen in Figure 5, this pipeline cycles the process of making 3d animated films and primarily shows rendering,texturing, and lighting to be in the production stage. This same stage spans for other animation techniques (behavioral,physics-based, voice puppetry) and thus the machine learning components will ideally take place before or in Research and Development. Pre-Production and Post-Production have too many stylistic dependence to be automated with Machine Learning. Rigging and VFX are just other stylistic parameters animators use to frame their presentation of a certain scene. These are better hand in a one-by-one basis on the animators due to the human workload being more reliant and cost-efficient (as mentioned before).

Figure 5:



III. Conclusion

In conclusion, these overviews of different AI technological procedures in the character animation shown the certain insights and following constraints around their performance compared to an animator's workload. As seen in Figure 5, the 3D production pipeline for creating a character have many shifting gears thus changes in these animation technologies will definitely" affect the schedules and budgets by shortening and bringing down costs." This incorporate of AI-driven applications will harness the use of such computer graphics in more movies and streaming shows that are to come. There will come to have more erry moments in film culture of character models to be resembling more life-like creatures from mimicking actual humans and other fantastic CG creatures.In the distant future, they can expand their decision-making can go toward Pre-Production stages like acting as script supervisors and in Post-Production stages like in (compositing) film-making.

By seeing these current AI techniques, one can surely see that they perform "best at filling in the blanks after a character already has been sketched out, either from an actual performance or by CGI artists. Another application could come into play in a movie like Guardians of the Galaxy Vol. 2, in which footage of current-day Kurt Russell and reference shots of his younger self were combined to "de-age" the character. AI "does sound magical," says Nichols, but "it's not exact data. It's interpreted data--interpreting the missing parts."" (Giardina 2017) Such processes are what most of these techniques in animation & rendering go for to overall incorporate/augment towards the animator's ability and not to replace their unique artistic endeavours whatsoever.

In this paper, Machine Learning in physics-based, and speech-based animations were the primary avenues of incorporation for quality assurement in the animated processes due to their coming incorporations and openness in experimenting with computer graphics or vision (respectively). Such instances can be found in Stephen Regelous, who got " a SciTech Academy Award recipient for the development of Massive, an AI-driven software that was first used to create the huge armies in Peter Jackson's The Lord of the Rings series" (Giardina 2017) With such a background, he also backs the optimization potential of AI when he said "You can shave off tens of millions of dollars from the budget of an animated feature. In a matter of seconds, AI could animate a character, a labor-intensive process for a CG animator." (Giardina 2017) Machine learning in Rendering and Behavioral Animation have many different project-specific mannerisms that can be daunting for a animated director to promote for his/her team.

Meeting the demands for the enormous amount of animated features in film and interactive mediums have resulted in computational or artistic-choice bottlenecks (as mentioned before). Thus these graphics techniques aim to help alleviate such bottlenecks. Gaining the confidence for these methods to be in put in practice and let alone in a disciplined fashion will be longer process to one might think. All in all, having machine learning and graphics being correlated has a promising back that these proposed methods will develop animation efficiency. After such corporate incorporations, AI animation techniques will have scalability in learning to master more nuanced topics like: geometric details, automation of an animator's stylistic tendencies (esp. hands and faces), and even the rules of cinematography. (Dinerstein and Cline 2006, 40)

IV. Reference List

Alfonsi, Benjamin. "Research applies AI to computer animation." *IEEE Intelligent Systems* 19, no. 4 (2004): 4-5. *General OneFile* (accessed April 1, 2019). http://link.galegroup.com.proxy.library.stonybrook.edu/apps/doc/A121747414/ITOF?u=sunysb&sid=ITOF&xid=d62b720c.

Dinerstein, Jonathan, Parris K Egbert, and David Cline. 2006. "Enhancing Computer Graphics through Machine Learning: a Survey." *The Visual Computer: International Journal of Computer Graphics*. ACM Digital Library. December. https://link-springer-com.proxy.library.stonybrook.edu/content/pdf/10.1007/s00371-006-0085-4.pdf.

Gao, Wen, Yiqiang Chen, Rui Wang, Shiguang Shan, and Dalong Jiang. "Learning and synthesizing MPEG-4 compatible 3-D face animation from video sequence." *IEEE Transactions on Circuits and Systems for Video Technology* 13, no. 11 (2003): 1119+. *General OneFile* (accessed April 1, 2019). http://link.galegroup.com.proxy.library.stonybrook.edu/apps/doc/A110738222/ITOF?u=sunysb&sid=ITOF&xid=53aff022.

Giardina, Carolyn. "Artificial intelligence: soon in a theater near You: script supervisors, film editors, CG artists and actors all had better look out: 'It's all over by 2045--we no longer are running the show'." *Hollywood Reporter*, August 23, 2017, 56+. *General OneFile* (accessed April 1, 2019). http://link.galegroup.com.proxy.library.stonybrook.edu/apps/doc/A504054378/ITOF?u=sunysb&sid=ITOF&xid=5edcd780.

Holden, Daniel, Jun Saito, and Taku Komura. 2016. "Neural Network Ambient Occlusion." *ACM Digital Library*. SIGGRAPH ASIA 2016 Technical Briefs. December 5. https://dl-acm-org.proxy.library.stonybrook.edu/citation.cfm?id=3005387.

Lau, Manfred, Ziv Bar-Joseph, and James Kuffner. 2009. "Modeling Spatial and Temporal Variation in Motion Data." *ACM Transactions on Graphics*. ACM Digital Library. December 19. https://dl-acm-org.proxy.library.stonybrook.edu/citation.cfm?id=1618517.

Taylor, Sarah, Taehwan Kim, Yisong Yue, and Moshe Mahler. 2017. "A Deep Learning Approach for Generalized Speech Animation." *ACM Digital Library*. ACM Transactions on Graphics. July. https://dl-acm-org.proxy.library.stonybrook.edu/citation.cfm?id=3073699.

Teicholz, Eric. "Computers: animation and rendering." *Progressive Architecture*, February 1990, 61. *General OneFile* (accessed April 1, 2019). http://link.galegroup.com.proxy.library.stonybrook.edu/apps/doc/A8360955/ITOF?u=sunysb&sid=ITOF&xid=1c1def28.

Vasilopoulos, Audrey. "Cartoon animation: computer animation tools are still playing 'catch-up' with traditional hand animation techniques." *Computer Graphics World*, January 1990, 77+. *General OneFile* (accessed April 1, 2019).

http://link.galegroup.com.proxy.library.stonybrook.edu/apps/doc/A8039144/ITOF?u=sunysb&sid=ITOF&xid=fd1aa0d8.

Zamuda, Ales, and Uros Mlakar. "Differential evolution control parameters study for self-adaptive triangular brushstrokes." *Informatica*, June 2015, 105+. *General OneFile* (accessed April 1, 2019). http://link.galegroup.com.proxy.library.stonybrook.edu/apps/doc/A459804218/ITOF?u=sunysb&sid=ITOF&xid=0c5ef2d3.