Graduate School of Advanced Science and Engineering
Waseda University

博　士　論　文　概　要
Doctoral Thesis Synopsis

論　　文　　題　　目
Thesis Theme

Research on Refinement and Real-time Editing of Expression for 3DCG Character Animation

申　　請　　者
(Applicant Name)

Takuya	KATO
加藤	卓哉

Department of Pure and Applied Physics, Research on Image Processing

December, 2018

Synopsis of the Doctoral Thesis Synopsis

Creating a 3DCG character animation requires a tremendous amount of work by hands. With its high potential on many applications not limited to entertainment, but also education, history and more, character animation had been used in a wide variety of fields.
When creating a character animation, there are mainly three layers of works that required to be done; Creating a basic model of the character, animate it with the parameters and edit to make it more realistic or aesthetical. Though the character we create might be identical, each layer of creation requires different kinds of the problem.
Creating a basic model of the character requires detailed sculpting. The quality of the final character fully depends on the modeling of a basic model. The dimensionality of the parameter also relies on the number of the vertices of the model, which ranges from a few thousand to a few million. Accordingly, even though the actual number of the parameters might be much smaller than the dimensionality of the data requires to be much large and complex.
After modeling of the basic model, animators set the motion parameters to generate the movement of characters. 3DCG characters are controlled by a simple parameter called 'rigs' that is parametrized for each pose or the muscle set defined by artists at the basic modeling phase. These rigs are mostly controlled by the parameters of 0 to 1 and set a number in the range to form a pose that the artists would wish the characters to take. Motion Capture, the tool to track the motion of the real human, will be used to transfer the motion to the character. While there are more practical tools used in this phase compared to the basic modeling phase, it still plays almost as the same amount of time as basic model creation. Even though the dimensions of the parameters are extremely small compared with basic modeling, effects of each rig parameters are much larger than each vertex of the model.
Then there is an editing phase. While we scan the actual faces. bodies or objects, setting all the parameters of the rigs for character on a motion capture, it is more than rare that each of these parameters can be directly applied to the character, and requires editing by hands. Especially for the motion, the small details changed in the editing phase plays even larger role comparing to the other phases, cannot be overlooked. Many hand works will be done to each of the rigs and the models to make the character have a better expressivity.

The problem that I have pointed out on each phase are completely different in many ways. Each artist has their specialty in solving each problem, which gives the answer to why each artist has a different specialty and choose them while learning to be a professional. One, and maybe the only, thing that each phase has in common is that we have a final version of each phase and that we have the data that was not used in the final production. Numerous and various reasons that the final version have been selected to be good, but we can absolutely say that there has been a difference between each of the versions, and the final version was regarded to be better. While it might be difficult for humans to even say which is the final version in a glance, it is not difficult to think that if we are able to capture such difference in each phase of production, it would be able for us to see what should have been done to make the model aesthetic. Such "garbage" data will never be been seen in the eyes of the public, while features can be considered to be an essential part of understanding the aesthetics in computer graphics to go beyond the uncanny valley.

This idea of defining aesthetics by the differences made in the process of the production is the most important idea of this dissertation. The expressivity cannot be defined in the same form of fashion, while the difference between the same topology of the data can always be defined as a refinement. In this dissertation, the refinement of character creation from roughly sculpted animation data to detailly sculpted animation is the key to understand the aesthetics of the 3DCG character and the effective way of using it will be the key for increasing the efficiency of creating characters. As this is to be the assumption, I have considered this assumption in the area of the facial expression sculpting and blendshape weight estimation in facial animation. Not only I considered how to synthesize the model automatically, but we also considered on how to control the character without corrupting the aesthetics by the help of the interface and the visualization of the controllers in order to support the artist. We considered the effectiveness in the challenging field of realtime character controls. This dissertation does not cover the whole area of the 3DCG to consider the assumption, though provide the basic idea of refining the characters aesthetics and empower the character animation productions.

In chapter 1, the introduction of the dissertation will be given. The introduction enters with some history of the character animation research as well as the background of how the 3DCG Computer Animation is created. It will also give you a brief overview of the challenges that this dissertation had been tackling and the contributions of my dissertation.

In chapter 2, the more detailed pipeline of the 3DCG character animation production will be given. Through the pipeline, you will be able to find how does the industry cooperate with the academia, what kind of the methods have been proposed compared and how does the research on my dissertation are unique from them.

In chapter 3, a method called "Character Transfer" that modifies the roughly created input blendshapes of arbitrary expressions to sculpt more specific individual expressions is introduce. Character Transfer sculpts the individualities by applying individual expressions extracted from a small number of training examples on segmented regions of a face. There are three contributions in this method. One is that this paper introduced the method to define the individual expressions as mapping. This allows us to extract individuality quantitatively so that it can be applicable to the other facial expressions. Secondly, this paper introduced a novel segmentation method that considers the geometry of the facial model and the facial expressions of the training examples. This allows the system to be effectively applicable even with the limited number of training examples. The third contribution is that this paper proposed a novel blending method of the individualities that avoids unnatural deformations caused by a naïve linear blending when applied on the input blendshape. By using this blending method it is possible to generate the individual expressions for arbitrary expressions. This paper show that Character Transfer can effectively modify the roughly created facial model with arbitrary expression with fully automatic algorithm other than sculpting training examples.

In chapter 4, the chapter proposes a method to automatically generate facial expression animation at singing from a few accessible inputs, which is singing voice information and song information. In this method, in addition to the singing voice information used in the speech animation generation method, this paper use song information to estimate expression parameter at singing and head rotation information in order to create a facial expression animation specialized at the time of singing. Paying attention to the fact that the behavior at the time of singing has a high dependence on the song, in addition to the volume and the acoustic feature amount at the time of singing highly correlated with the head rotation information at the time of singing, information on the rhythm and melody are adopted. Furthermore, mouth shape information which changes according to lyrics is estimated from singing voice and adopted as input data. By making the learning of these data into a deep learning model which take time series into consideration, the method made it possible to generate facial expression animation at singing which was difficult only with singing voice information. In this paper, in addition to a quantitative evaluation of these results with measurement data, this paper verify the naturalness of the generation result and whether animation unique to each singer can be learned by subjective evaluation experiment. This will discuss the validity of the input parameters and the stability of the generation result.

In chapter 5, the dissertation presents a novel dance authoring interface called DanceDJ based on such a hypothesis. By mimicking the interaction of the musical DJ, the system allows users to control dance motion more intuitively and create high-quality dance motion in real-time. By implementing the synchronize button that synchronizes the dance motion with the music played by the other electronic instruments, the user can match the beat of the dance motion to that of the music. In contrast to the music DJ, the connectivity and beat of the dance motion tend to be more abstract than that of the music, which affects the usability when connecting the dance motion. To support the users, this work have implemented a novel feature to estimate the connectivity of the dance motions. The system automatically calculates the beat of the dance by using the motion intensity to estimate the probability of the frame-wise dance connectivity, which represents how well the beat of dance and music matches together. This feature helps the users to achieve a DJ-like experience when interactively creating dance motion and allows users to create, as well as realizing our hypothesis of the beat and correlation with music. This work have produced a fully usable system for real-time dance authoring. Experimental results show that high-quality dance motion can be synthesized in real-time using the proposed DJ interface. This work has performed quantitative user studies to support the usability of the system. For the user, I have evaluated how well users have been able to create high-quality dance motion intuitively. For the audience, I have evaluated how plausible the created dance motions are by looking at the dance motion results our system has created in real-time. I have also conducted tests of using our system in real-world live performances. I have performed quantitative user studies to support the usability of the system.

In the last chapter, I will be concluding the dissertation with some problems that require to be solved in the future with some problems to be expected in near future.

早稲田大学　博士（工学）　学位申請　研究業績書

(List of research achievements for application of doctorate (Dr. of Engineering), Waseda University)

氏　名(Full Name)　Takuya Kato　　　　印(seal or signature　　　　)

（As of February, 2019）

種 類 別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）(theme, journal name, date & year of publication, name of authors inc. yourself)
論文	
○	加藤卓哉，深山覚，中野　倫靖，後藤真孝，森島繁生，“歌声と楽曲構造を入力とした歌唱時の表情アニメーション自動生成手法”，画像電子学会誌, April, 2019.
○	“DanceDJ: A 3D Dance Animation Authoring System for Live Performance”, Lecture Notes in Computer Science, vol 10714. Springer, Cham, Dec. 2018, Naoya Iwamoto*, Takuya Kato*, Hubert P. H. Shum, Ryo Kakitsuka, Kenta Hara, Shigeo Morishima. (* equally contributed)
	他 1 件
国際会議	
○ 査読あり	“DanceDJ: A 3D Dance Animation Authoring System for Live Performance”, Advances in Computer Entertainment Conference - ACE 2017, December 2017, Naoya Iwamoto*, Takuya Kato*, Hubert P. H. Shum, Ryo Kakitsuka, Kenta Hara, Shigeo Morishima (* equally contributed)
○ 査読あり	“Character Transfer: Example-based individuality retargeting for facial animation”, International Conference in Central Europe onComputerGraphics,VisualizationandComputerVision2014(WSCG2014), WSCG 2014, Full Papers Proceedings, 121-129, June 2014, Takuya Kato, Shunsuke Saito, Masahide Kawai, Tomoyori Iwao, AkinobuMaejima, Shigeo Morishima.
○ 査読あり	"Example-Based Blendshape Sculpting With Expression Individuality", ACM SIGGRAPH 2014 Poster, 7, Vancouver, 2014.08.10-14. Takuya Kato, Shunsuke Saito, Masahide Kawai, Tomoyori Iwao, Akinobu Maejima and Shigeo Morishima.
	口頭発表　他 3 件 ポスター　他 12 件

早稲田大学　博士（工学）　学位申請　研究業績書

(List of research achievements for application of doctorate (Dr. of Engineering), Waseda University)

種 類 別 By Type	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）(theme, journal name, date & year of publication, name of authors inc. yourself)
国内会議	
査読あり	"DanceDJ: ライブパフォーマ ンスを実現する実時間ダンス生成システム" WISS2017, 山梨, 2017.12.6 - 8., 岩本尚也*, 加藤卓哉*, 原健太, 柿塚亮, 森島繁生. *共同主著
査読あり	"キャラクタの個性的な表情特徴を反映した表情モデリング法の 提案." フォーラム顔学 2015, 名古屋, 2015.09.12-13. 加藤卓哉, 森島繁生.
査読あり	"Character Transfer キャラクタ固有の表情特徴を考慮した顔ア ニメーション生成手法." Computer Entertainment Developers Conference 2014 (CEDEC2014), 横浜, 2014.09.02-04. 加藤卓哉, 森島繁生.
	他 5 件
講演	"Dance DJ: ライブパフォーマンスのためのダンス動作ミックスシステム", WISS2016, 滋賀, 2016.12.14 – 17 岩本尚也, 加藤卓哉 , 原健太, 柿塚亮, 森島繁生.
	"キャラクタに固有な表情変化の特徴を反映したキーシェイプ自動生 成手法の提案." 情報処理学会 第 76 回全国大会, 4ZC-3, 東京電機 大学, 2014.03.12. 学生奨励賞 加藤卓哉, 斉藤隼介, 川井正英, 岩尾知頼, 前島謙宣 森島繁生.
	"キャラクタ特有の特徴再現を考慮したリアルな表情リターゲッティング手法の提案." 情報処理学会 グラフィックスと CAD 研究会第 154 回研究発表会, 15, 理化学研究所, 2014.02.2.1, 加藤卓哉, 川井正英, 斉藤隼介, 岩尾知頼, 前島謙宣 森島繁生.
	. "キャラクタ固有の 表情特徴を考慮した顔アニメーション生成手法." Visual Computing/GCAD 合同シンポ ジウム 2014, 37, 東京, 2014.06.29-30., 加藤卓哉, 斉藤隼介, 川井正英, 岩尾知頼, 前島謙宣 森島繁生.
	他 9 件