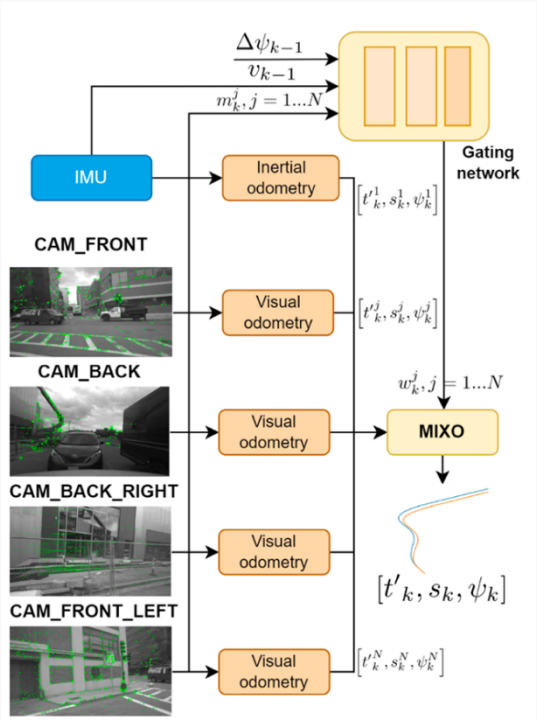
IEEE CONSUMER TECHNOLOGY SOCIETY (CTSOC)

ISSUE: MARCH2023

CTSOC-NCT NEWS ON CONSUMER TECHNOLOGY



The Mixture of eXpert Odometry (MIXO) architecture, a novel way to dynamically weight the contribution of each camera using a data-driven approach. (MIXO) technique obtains a loosely coupled joint estimate for the vehicle trajectory, building upon individual odometry measurements that are less computationally demanding than solving a high-dimensional joint optimization problem. Their proposed method can also be extended to fuse camera informa- tion with other sensors (e.g., inertial). Keypoint detection and matching is performed on each individual camera stream. Yaw rate estimates provided by each expert (camera) are weighted by a gating network, which takes as input the yaw rate and velocity at the previous frames, as well as the number of matched keypoints in the current frame. The gating network is composed by two fully connected layers, with dropout, followed by a softmax layer.

IEEE CTSoc CONSUMER TECHNOLOGY SOCIETY

TABLE OF CONTENTS

EDITOR'S NOTE

On behalf of the Editorial Board of IEEE CTSoc News on Consumer Technology (NCT) editor-in-chief Wen-Huang Cheng, and fellow co-editors, Yafei Hou, Luca Romeo, Jianlong Fu, and Chuan-Ju Wang, I am pleased to introduce the March 2023 issue of the News on Consumer Technology (NCT).

For this issue, we begin with a cover story of MIXO, a Mixture of Experts-based visual odometry approach to estimate vehicle trajectory from multiple cameras equipped in autonomous vehicles. The MIXO is a novel method that dynamically weighs the contribution of each camera using a data-driven approach resulting in a more robust and accurate technique as compared to single camera methods. Moreover, it is a lightweight module that is easily implemented on top of any existing visual odometry algorithm with potential for expansion in various ways.

This is followed by an interview with Mr. Eric Tsai, the Chief Technology Officer of KKCompany Technologies, a leading technology group in Asia based in Taiwan. Mr. Tsai shares his experiences starting from his IT background as the first employee of the company during its founding in 1999, till his transition into a business manager and various leadership roles in the company, including his vision for the company in pushing forward streaming technology.

Finally, this issue ends with a feature article by Prof. Dr. Shih-Wei Sun and Ms. Yu-Hsuan Lo, from the Department of New Media Art, Taipei National University of the Arts, Taiwan. This article introduces the DeepHair project and discusses the potential for positive values of DeepFake technology in a time where the general public is increasingly concerned with the technology's negative implications.

Have a nice read!

Yuen Peng Loh
Editor of NCT

ARTICLE TITLE
MIXO: Mixture of Experts-based Visual Odometry for Multicamera Autonomous Systems

AUTHOR(S)
Lia Morra; Andrea Biondo; Nicola Poerio; Fabrizio Lamberti

JOURNAL TITLE
IEEE Transactions on Consumer Electronics

JOURNAL VOLUME AND ISSUE
Early access

DATE OF THE ARTICLE
January 2023

DOI
10.1109/TCE.2023.3238655

Visual-inertial odometry is a fundamental technology exploited by autonomous vehicles and mobile robots to determine their position in unknown environments. Indeed, many techniques have been proposed for monocular and stereo cameras. State-of-the-art autonomous vehicles, however, are equipped with multiple cameras covering the entire vehicle environment. In this paper the authors present a MIxture of eXperts Odometry (MIXO), a data-driven, machine learning-based technique that loosely combines odometry outputs from multiple cameras to obtain a more accurate and robust global estimate. It stems from the intuition that each camera provides an optimal vantage point in specific driving scenarios. In MIXO, each camera (or expert) is individually processed by a state-of-the-art visual odometry algorithm (e.g., ORB-SLAM2). Then, the odometry estimates are mixed by a gating network, which selects the locally optimal experts in the current operational conditions and weights their contributions accordingly. MIXO is a lightweight module that can be easily implemented on top of any visual odometry algorithm. Experimental results on real-life data from autonomous vehicles show that MIXO achieves more robust and accurate results than any single camera, reducing the absolute rotational and translation error by 38% and 15%, respectively.

INTERVIEW WITH ERIC TSAI, CHIEF TECHNOLOGY OFFICER OF KKCOMPANY TECHNOLOGIES

Eric Tsai

Eric Tsai, the Chief Technology Officer of KKCompany, holds the honor of being the first employee and has been with the company since 1999. Over the past 20 years, you have held many positions; what are your reflections on your career so far?

I joined the company early on as its first employee when it was founded in 1999. The following year, our company launched KKMAN, which had a 22% market share of Taiwan's web browsers. Later, in response to the rise of mobile devices, we created a second growth momentum by developing applications such as mobile ringtones and games. In 2005, following the development of broadband and the rise of digital music, we invested in multimedia services and launched KKBOX, the world's first legal music streaming platform, and then expanded into markets overseas, including Hong Kong, Singapore, Malaysia, and Japan. In 2016, the company became an enterprise group and developed multiple service brands such as KKTV and KKStream, as I started to serve as the President until 2022, when I returned to KKCompany Technologies (KKCompany) as the CTO.

Over the years, I have held various positions within KKCompany: KKBOX, KKTV, and KKStream; from R&D to business management, and my

mindset has developed along this journey. I think the biggest realization is that "technology development excellence does not equal commercial success." We have to stand in the customer's shoes in order to resolve problems.

With my technology-based background, I used to assume that as long as the application and system were well-designed, the product would sell well. But this is not the case at all. If you promote the wrong information and story to the market, you will not succeed even if you have the best technology.

The first time I realized this was during a presentation with a customer. After listening to the presentation, the customer still didn't understand the highlights of the technology and why they should apply our solutions. After a few trials and errors, I gradually learned to put myself in others' shoes, which is not a trait all engineers possess. Until now, I still constantly remind myself that I have to balance the scale from time to time. Every experience is an interesting and challenging one that keeps opening up new horizons for me.

Since its establishment, KKCompany has achieved technological advances such as music streaming, multimedia technology, and even cloud intelligence; how do you see the changes and opportunities from the company's perspective?

Over the past 20 years, we have seen many trendy technology buzzwords that we all thought were transformative at first eventually fade into oblivion. Only one out of every ten technology buzzwords ends up making an impact on the market trend.

In the early days, when we saw a trending technology term, we rushed to activate product research and development. However, once we realized we were headed in the wrong direction, it became challenging to change course. Like the popular Generative AI, there are indeed many applications that can be explored, but perhaps five years later, when we look back at the development of AI, we will realize that the current AIs are still at a preliminary stage; our norms five years later may seem impossible to us right now. I am also not convinced that everything will be overturned by a single technology altogether.

Therefore, we encourage teams to try out new technologies by being nimble and smart: experimenting daringly with agile development processes and researching technologies from a multi-dimensional perspective, such as organizing hackathon competitions to challenge technologies in different fields and opening up to proposals from everyone to stimulate innovation in software applications.

How does KKCompany innovate as new technologies and markets emerge? What is the process?

We adopt various strategies for internal innovation. In the early days, it was more often a "top-down" approach. In the past few years, our colleagues have become well-trained, and the ideas they propose are mature enough that they do not differ much from reality in the market in terms of feasibility. The most important aspect to think about is the business and profit model behind the proposal. Therefore, after each idea has taken shape, we have to survey different units within the group to ensure that we can eventually produce a solid business plan.

To develop a new feature, you need to determine what you want to achieve in the end; you cannot just rely on intuition to make decisions, for example:

When KKBOX first ventured into music streaming services, the music would play within a second after pressing the play button in a good network environment (the office). So the team wanted to challenge the Hi-Resolution audio function right from the nascent stages of the development.

One day, a colleague went home to listen to music using a dial-up network but discovered that it took a long time for the music to start playing, resulting in an awful user experience.

This — users' bad experience of the Hi-Res

audio function — is something that the developers could not have imagined in the office environment, and without verification from different angles, KKBOX may no longer exist today.

Similar experiences also helped me to think about things from different aspects when I was in charge of the company's operation later on. If you think one step ahead of others, you can potentially spend less effort later on.

You have transitioned from an IT person to a business manager. What advice would you give to workers who also have an engineering background and want to enter the technology industry from your past experience in terms of mindset change, career development, and management?

The most important thing is to have "empathy" when doing everything: we must remember to zoom backward and outward and gain perspective to observe aspects we might not have previously thought of. Furthermore, try not to think in a small silo — engage more with colleagues, discuss, and learn from others whenever you don't understand.

Of course, it is easier said than done, but it will naturally become a habit. Then, when planning things, the opportunity costs of doing one thing will naturally arise in your mind, and you will also consider whether there is an opportunity to integrate other great things and boost the results of each project.

After getting used to this model, when you encounter something you are not sure of, you can turn it into a set of implementation guidelines.

In my early career, I didn't like to do annual planning. My first reaction was always, "What's the point of planning when things are going to change anyway?". I'm sure many people with engineering backgrounds share this mentality.

Later on, I realized that planning is a concept, and the general direction can, of course, be modified, but if it is poles apart from the initial aim, it won't work in the end; without an outline at the beginning, the team will be like a headless chicken.

Therefore, I would advise those who want to enter the tech industry with an engineering background to set a clear orientation at the beginning and communicate with their team regularly to make sure that everyone's understanding is consistent. This is one of the most difficult things to do when leading engineers, and this is also what we are striving to improve on.

Technical people are good at solving specific problems; business people are good at finding operational balance.

The transition between the two identities is akin to frequent collisions. From the cases I have observed in the past, many people will pass the buck to other colleagues when they encounter a bottleneck, which is a pity because they miss an opportunity to hone their skills.

However, this doesn't imply that leaders should bear the burden of everything, but rather that they should broaden their perspective and develop the skills to lead people on a day-to-day basis in their work instead of waiting to develop leadership skills until they've already assumed the position as a leader.

To maximize value in the future, it is essential to foster cross-disciplinary collaboration and integrate diverse fields of study.

When I interview engineers, I often inquire about their career aspirations. Many express a desire to lead their teams. Then, I ask, 'Aside from earning more money, do you have any other motivation for the leadership position?' A lot of individuals are unable to answer this question.

Leadership involves complementing a team, amplifying its capabilities, and creating greater value. Although challenging, this prompts individuals to reflect on the competencies they must acquire and the trade-offs they may need to make.

KKCompany has been working closely with academic institutions since 2007; what was the initial opportunity, and what are the concrete results and cases of collaboration of the

present day? Has the company identified any applications that can be commercialized in practice?

The earliest collaboration came from campus recruitment in 2007, when the KKBOX brand was still not well-known yet. In order to recruit better talents, I started to hold sharing sessions on campus to foster linkages with academia. Through this, I discovered that many school labs were already doing research on machine learning, which led to "music emotion analysis" related applications.

A song can be classified according to various indicators, such as the music style, lyrics, and emotion. These categories can help streaming platforms provide more personalized song recommendations and curated playlists. In the past, we relied on manual tagging for classification. At that time, using AI for filtering and classification was considered advanced technology.

I think industry-academia collaboration is a good entry point so that the research done by the university is no longer just academic work to be delivered for graduation but also a profound contribution to the industry.

The first generation of KKBOX's recommendation engine was born in industry-academia cooperation with National Sun Yat-sen University. Since then, we have been cooperating with National Chengchi University, National Taiwan University, and Academia Sinica and have gradually worked out a model of cooperation between school and industry.

What are the areas where KKCompany would like to collaborate with academia in the future?

Our next objectives are video and music streaming, AI, and other multimedia applications. For instance, as more surveillance cameras are installed on public transportation, we need to enhance our image recognition and detection capabilities. Although this is not our core business today, we have a robust streaming system that can be integrated with cross-disciplinary applications.

In addition, there are potential applications for healthcare institutions where caregiving requires long-term video recording, and those data need to be stored properly. Our company is not an expert in the medical field, but we offer advanced streaming technology. This kind of endeavor of a cross-industry, experimental nature, or products in the POC (Proof of Concept) stage can be achieved the fastest through industry-academia cooperation.

DEEPHAIR: FROM AN ARTWORK TO AN INTERACTIVE APPLICATION

Yu-Hsuan Lo
ms26703369@gmail.com
Yu-Hsuan Lo is a master student at the Department of New Media Art, Taipei National University of the Arts, Taipei, Taiwan. Her research interests are deep learning applications, computer vision applications, and VR/AR applications.

Shih-Wei Sun , Ph. D.
swsun@newmedia.tnua.edu.tw
Shih-Wei Sun is a Professor at Department of New Media Art, Taipei National University of the Arts, Taiwan. He received his Ph.D. in Electrical Engineering from the National Central University, Taiwan. His research interests are computer vision, sensor applications for mobile devices, and applications for human-computer interface.

Recent development of AI technology, especially generative models such as DeepFake and ChatGPT, has garnered concerns in the general public as well as research and governance communities due to the high risk of misuse and negative consequences garnered from unregulated access by unsavory individuals. However, AI technology in itself is a product of technological progress that is not innately harmful. This article introduces the DeepHair project and discusses the positive value that can be brought by DeepFake technology, despite the negative associations it is commonly known for.

Introduction on DeepHair

The project DeepHair [1] is a hairstyle preview system that is proposed based on the DeepFaceLab [2], a leading software for creating DeepFakes. The proposed interactive hairstyle preview system allows a user to stand in front of a camera with an interactive touchable instruction panel to choose the preferred hairstyle video captured from the hairstyle models, as shown in Fig. 1. A hairstyle preview video will be generated based on the selected hairstyle via the DeepFake method. Compared with the conventional image-based hairstyle preview systems, e.g, GAN-based method [3], and StarGAN method [4], the proposed DeepHair allows users to view the generated hairstyle video with more dynamic coverage from a variety of view angles.

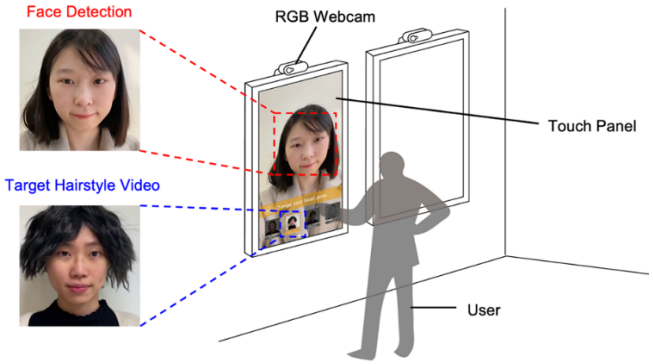


Figure 1: The system diagram of DeepHair [1].

The Artwork <To Be>

Prior to the system development, the project has been explored in a more artistic avenue. The

predecessor of the DeepHair project [1] is the artwork entitled, <To Be, Yu-Hsuan Lo, 2021> as shown Fig. 2. In the artwork, the artist, Yu-Hsuan used a camera to record a short video of herself using OpenCV to detect the facial parts from the frames, and DeepFaceLab [2] was applied to integrate the facial parts into the target videos. As shown in Fig. 2 (a), the target videos consist of four different characters[1]: a boy (right part[2]) , a little girl (upper part[3]) , a teenage girl (left part[4]) , and an office lady (lower part[5]) . The characters represent four different identities that stayed in Yu-Hsuan's mind. Furthermore, on social media platforms, Yu-Hsuan assumed the roles of the four characters with different user IDs. In this artwork, the videos played on the four cubes in Fig. 2 (a) are generated from DeepFake methods. In the central parts, the virtual human subject represents the real Yu-Hsuan, and the four DeepFaked videos are the characters played Yu-Hsuan as the "faked" characters in the social media platforms and game platforms with the respective identities. The videos were generated by the "face swapping" process in DeepFaceLab [2] and displayed on the cubes as illustrated in Fig. 2 (b). In the artwork <To Be>, Yu-Hsuan discussed an in-between relationship between an inside character (the central virtual figure in Fig. 2(a)) and the expected characters in the social world (the images in Fig. 2(b) projected outwardly by the four cubes) of the same person in the real physical world.

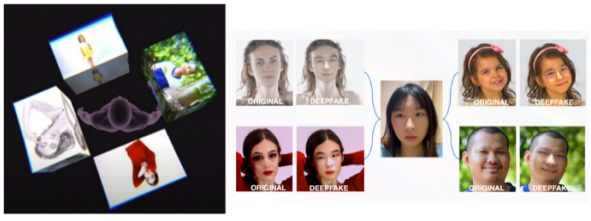


Figure 2: The artwork <To Be, Yu-Hsuan Lo, 2021>: (a) The artwork with four different cubes and (b) Yu-Hsuan Lo's facial featured replaces those of the characters in four different videos based on DeepFaceLab [2].

The Transition to an Interactive Application: A Negative or Positive?

During the creation of the artwork <To Be, Yu-Hsuan Lo, 2021>, the news broke the story of a Taiwanese Youtuber "Xiao Yu" who DeepFaked videos [5], specifically using DeepFake technology to impose celebrities' faces over actors' in pornographic videos. Due to this incident, DeepFake technology is widely known as an AI application with a negative undertone. In addition, DeepFaked videos shared over the social media raised issues of invasion of privacy and disinformation [6]. Furthermore, ChatGPT [7] in the text retrieval domain has been developed to use a machine-trained model to provide conversational applications as dialogues from a user to a cloud server. The generated texts are as natural as the sayings of a human being. The situation is similar to the videos generated by the DeepFake [2] method to be viewed by audiences. Not only texts, but also the videos, are hard to be distinguished by the human users whether they are real texts/videos or faked/generated texts/videos. When the technologies are applied in negative ways, this confusing property makes the situation worse. In order to detect such fake texts, Mitchell et al. developed "Detect GPT [8]" whereas, Ciftci et al. developed "FakeCatcher" [9] to detect fake videos. More and more effort has been paid by researchers to detect faked media contents [8] [9] so as to suppress the situation from getting worse.

Despite the negative implication on the DeepFake name, the technology is in fact neutral as demonstrated with the development of the artwork <To Be, Yu-Hsuan Lo, 2021> where it is used as an abstract exploration of the self and the outward characters portrayed to the social world. In particular, we applied light fitting to estimate the features from various angles of views in DeepFake to generate a more natural face swapping effect from the artist Yu-Hsuan to the target video belonging to a character with his/her identity to generate a virtual character in a virtual space. In this context, the DeepFake approach is merely a tool of artistic expression of a user and not an inherently harmful technology.

So, to move on from the negative, we focus on the capabilities and potentials brought upon by the artwork. Since the face swapping from the user (the artist Yu-Hsuan) to a character can be displayed

[1] Videos were obtained from the copyright free website, Pixabay.
[2] https://pixabay.com/videos/man-smiling-farmer-happy-young-76888/
[3] https://pixabay.com/videos/kids-dancing-outfits-new-year-12266/
[4] https://pixabay.com/videos/weapons-pistol-depression-stress-45282/
[5] https://pixabay.com/videos/girl-posing-model-portrait-lovely-46948/

naturally, as shown in Fig. 2, such facial swapping has great potential to be developed into applications. We were thinking about: "Is there an application that needs to swap the facial part to a target video?" Thus, after our brainstorming, we came to the scenario of a hair salon. To have a haircut in a hair salon, it is not easy for a customer to estimate the final result of a haircut, and the communication between a customer to a hair stylist is often time consuming if not inaccurate. During a haircut, the customer may be in a constantly anxious state throughout the whole session.

Therefore, we proposed DeepHair [1], a Deepfake technology for swapping the facial part of a customer in a salon to a target video with a hairstyle model. In this circumstances, issues of invasion of privacy and disinformation are not touched, and the generated video is a demonstration of a positive implication, that is as a medium for a customer and a hair stylist to have common understanding and effective discussions regarding the hairstyles. The obtained hairstyle preview video can release the pressure of misunderstanding from both the customer and the hair stylist in a salon and promote favorable communications. Hence, the proposed DeepHair [1] system is an exhibition of a positive use case of the DeepFake technology.

The DeepHair System

In the developed DeepHair [1] system, a customer in a hair salon can select a video with a hairstyle model (the video with a human subject at the bottom-left of Fig. 3) from our system. Next, a customer can use a camera to record a small period of video with her face (the top-left of Fig. 3) as the input data. The hairstyle is then transferred by the adapted DeepFaceLab [2], on to each frame of the customer video. An example of the target losses measured among different epochs are shown on the top-right of Fig. 3, and the corresponding generated deepfake frames are shown on the bottom-right of Fig. 3.

As a prototype, a demo system of DeepHair [1] was developed and showcased to the authors and participants in ACM Multimedia Asia 2022, so as to provide a practical interactive user experience. Figure 4(a) shows the process where a user can use her finger to select the target video with a hairstyle. When recording a short video as the input to the system, a user is asked to rotate her head to different directions, as shown in Fig. 4(c)-(d). Finally, the face swapping results from the input short video to the hairstyle model video are obtained, and the hairstyle preview from different angles can be observed as shown in Fig. 5. This hairstyle preview video can then be used a point of discussion between a customer and a hair stylist in a hair salon.

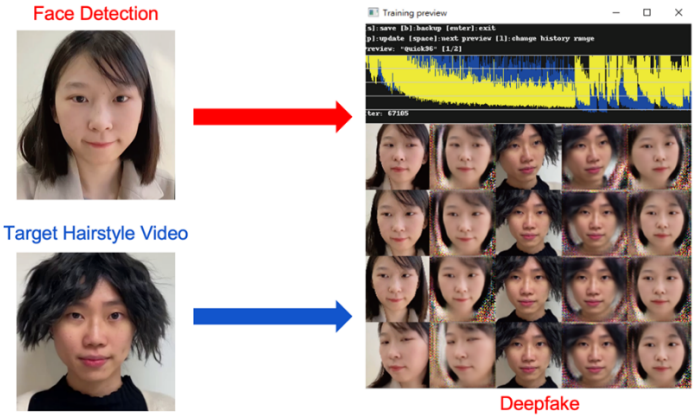


Figure 3: Hairstyle transfer by adopting DeepFaceLab [1] [2].

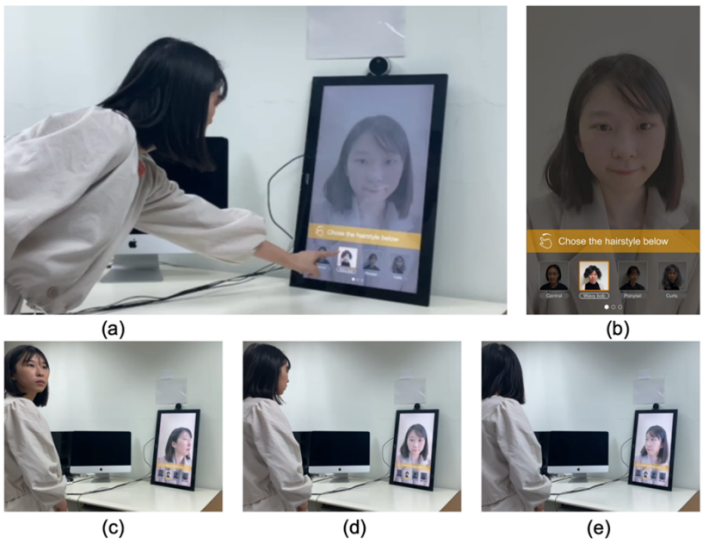


Figure 4: System interaction, (a) choosing a target hairstyle video, (b) the screen shot of the system UI, (c) rotating the head to show the left face, (d) rotating the head to show the middle face, (e) rotating the head to show the right face [1].

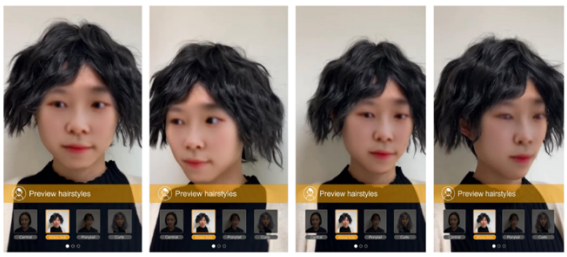


Figure 5: Hairstyle previewing result from different viewing angles in a generated DeepFake video [1].

Now and the Future

Currently, DeepHair [1] is a pilot study of integrating AI technologies to create artworks as well as practical positive uses in the future. In terms of technical implementations, the adoption of DeepFace Lab [2] for the DeepHair [1] still requires significant computational power in order to generate the DeepFake preview video, from the current 15 minutes to a more manageable period of time. Cloud computing or distributed DeepFake processing approaches can be developed in the future to support a system in this aspect. In addition, the DeepHair [1], actually performs 2D frame-to-frame swapping, therefore, 3D processing is expected to be more effective. By scanning the 3D point clouds of a user's face, we expect the outcome of the face to be more natural, and the 3D model generation may allow users to interact with the generated 3D content more efficiently. Furthermore, the proposed hairstyle preview system [1] can also be deployed into web services in the future. For example, the virtual try-on glasses of OWNDAYS [10] (a Japanese eyeglass company), and the virtual makeup system of L'Oréal Paris [11]. The use of built-in camera of the computers of the users to provide an AR-based user experience of virtual-on holds great potential. In the future, we expect a hairstyle preview system can be used in hair salons to be used on portable tablets, pads, and even smartphones as a testament that DeepFake technology can be used for good.

References

[1] Lo, Y. H., & Sun, S. W. (2022) DeepHair: a DeepFake-based Hairstyle Preview System. In Proceedings of the 4th ACM International Conference on Multimedia in Asia, 41, 1 - 2.
[2] Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., Jiang, J., RP, L., Zhang, S., Wu, P., & Zhang, W. (2020). DeepFaceLab: Integrated, flexible and extensible face-swapping framework. arXiv preprint arXiv:2005.05535.
[3] Zhu, P., Abdal, R., Femiani, J., & Wonka, P. (2021). Barbershop: GAN-based image compositing using segmentation masks. ACM Transactions on Graphics (TOG), 40(6), 1-13.
[4] Nakamae, Y., Wang, X., & Yamasaki, T. (2020). Recommendations for Attractive Hairstyles. In Proceedings of the 2020 Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia, 19-24.
[5] Benack, E. (2021). Taiwanese Youtuber uses deepfake tech to make celebrity porn, Radio Taiwan International. Retrieved February 28, 2023, from https://en.rti.org.tw/news/view/id/2006342
[6] Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107, 1753.
[7] OpenAI. (2022). ChatGPT: Optimizing Language Models for Dialogue. OpenAI. Retrieved February 28, 2023, from https://openai.com/blog/chatgpt
[8] Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. arXiv preprint arXiv:2301.11305.
[9] Ciftci, U. A., Demir, I., & Yin, L. (2020). Fakecatcher: Detection of synthetic portrait videos using biological signals. IEEE transactions on pattern analysis and machine intelligence. (Early Access)
[10] OWNDAYS. (n.d.) Virtual Try-On. Retrieved February 28, 2023 from https://www.owndays.com/vn/en/try-on
[11] L'Oréal Paris. (n.d.) Virtual Makeup Try-On. Retrieved February 28, 2023 from https://www.lorealparisusa.com/virtual-try-on-makeup