Generating Personalized Virtual Agent in Speech Dialogue System for People with Dementia

Shota Nakatani[1(✉)], Sachio Saiki[1], Masahide Nakamura[1], and Kiyoshi Yasuda[2]

[1] Graduate School of System Informatics Kobe University,
1-1 Rokkodai, Nada, Kobe, Japan
shota-n@ws.cs.kobe-u.ac.jp, sachio@carp.kobe-u.ac.jp,
masa-n@cs.kobe-u.ac.jp
[2] Chiba Rosai Hospital, 2-16 Tatsumidai-higashi, Ichihara, Japan
fwkk5911@mb.infoweb.ne.jp

Abstract. Our research group has been studying a speech communication system with a virtual agent (VA), to support person-centered care (PCC) of people with dementia (PWD). The current system uses the 3D model based on an unreal character for the VA. Because the unfamiliar appearance is to be a mental obstacle to PWD, PWD hardly accept advice and which causes a limitation in the care effects. In this paper, we develop a novel system that dynamically creates a VA based on a given facial image of real person. The proposed system constructs a three-dimensional model based on facial landmarks within the image. It then stretches and transforms some portions of the 3D model to generate facial expressions. From just a given picture, the proposed system easily generates a communication agent familiar with individual PWD. Hence, it can implement (virtual, but effective) conversations with familiar partners. We implement the prototype based on the proposed system and conduct the experiment targeting to the elderly.

Keywords: Virtual agent · Home elderly care · Person-centered care

1 Introduction

Japan is facing a hyper-aging society. The number of people with dementia (PWD) will reach 7 million in 2025, where one-fifth of five elderly people in Japan will suffer from dementia [1]. Hence, a care and a support for PWD are socially needed.

The person-centered care (PCC) is an ideal care for PWD, which monitors and understand individual circumstances, and plans and executes optimized care. The PCC is different from the conventional care and needed to watch care subject sufficiently, which poses heavy physical and mental burden on family and caregivers. In practice, however, the PCC completely relies on human effort.

V. G. Duffy (Ed.): DHM 2018, LNCS 10917, pp. 326–337, 2018.
https://doi.org/10.1007/978-3-319-91397-1_27

To cope with the problem, our research group has been studying a PCC support system for PWD, exploiting the latest IoT and cloud technologies [12]. Gathering and analyzing sensor data from the home of PWD, the system understands activities and contexts of PWD at home [11]. The system then generates dialogues, and talks to PWD through the virtual agent technology [9]. The virtual agent (VA) is a human-looking animated chat-bot program operated on a lap-top PC. Using the speech recognition and synthesis technologies, a PWD can communicate with the VA as if (s)he talks to a human partner.

However, the current system uses an unreal and artificial avatar for the VA. In social psychology, the faces have a large effect on partner in communication between humans. It is considered the same thing in communication with the VA, since you communicate. Hence, because of unacceptance of the VA's appearance in current system, the care and advices from unfamiliar avatar does not motivate the PWD very well. It poses limitations on concentration and engagement of the PCC.

This paper develops a new technology that introduces more familiar agent as the VA, in order to achieve more concentrated and effective care within the PCC system. We use just a picture of face in method to generate the VA. As a result, the system can easily display the avatar of the person as the VA, who is familiar with PWD or influential in PWD.

We generate a three-dimensional model from a given picture of a face of a familiar person of a PWD, and integrate the 3D facial model as the VA of the PCC support system. More specifically, for a given picture of a face, the system extracts landmark points of the face using a face recognition algorithm. Based on the landmark points, the system then stretches and transforms some portions of the 3D model, to dynamically generate facial expressions. Also, for a given voice data, the system synchronizes the lip motions, as if the 3D model speaks to the PWD. By integrating the generated model to the PCC system, we expect to relieve PWD of resistance to care and communication in an existing PCC support system.

From just a given picture, the proposed method can dynamically generate a communication agent according to preference of individual PWD. Hence, it can implement (virtual, but effective) conversations with familiar partners of PWD, such as close relatives, child, close friends and so on. As a result, the proposed system contributes to the implementation of more effective PCC support system.

2 Preliminary

2.1 Person-Centered Care Support System with Virtual Agent

Our research group has been studying how the ICT can support a person with dementia (PWD) at home. The concept of person-centered care (PCC) defines an ideal care for PWD, which monitors and understand individual circumstances, and plans and executes optimized care. Our current aim is to provide person-centered communication for PWD using the virtual agent (VA) technology. The VA is a human-looking animated chat-bot program operated on a PC. Using the

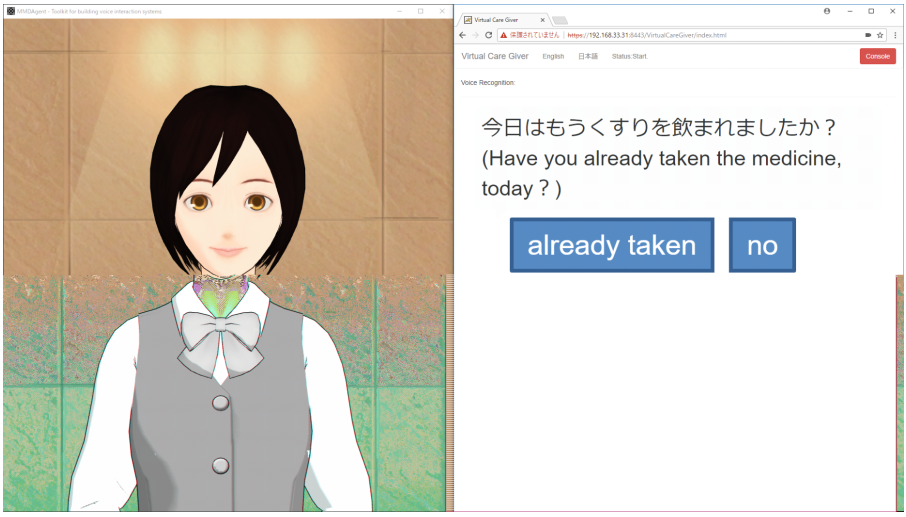

Fig. 1. VirtualCareGiver

speech recognition and synthesis technologies, PWD can communicate with the VA as if (s)he talks to a human partner.

In our previous research, we have developed a system, called Virtual Care Giver (VCG), using the VA [12]. Figure 1 shows the image of VCG. VCG was designed to be able to cooperate with Web services, to integrate IoT, smart home and cloud. Because of this design, VCG can generate personalized cares and conversations from activities and contexts estimated based on sensor data gathered from the home of a PWD. VCG then can provide a care and conversation for the PWD via VA, which supports the PCC. Delegating the communication care with VCG, a human caregiver can concentrate on human-centric tasks that cannot be done with ICT.

Our preliminary experiment shows that VCG can achieve useful context-aware care for PWD, including, daily greeting, schedule reminder, and prompting medication.

However, there is also a limitation caused by the looking of VA. Currently, the VA of VCG is implemented by MMDAgent [7], which displays a 3D animation character like Fig. 1. So, the care and advices from unfamiliar avatar do not always motivate the PWD very well. This poses limitations on concentration and engagement of the PCC.

2.2 Face Recognition in Cognitive Computing

Cognitive computing is a computing paradigm where a system imitates the human brain and learns itself to derive answers. It often refers to the emerging technologies that can analyze non-numerical data, such as language, picture and voice, which were difficult to understand by the conventional computing.

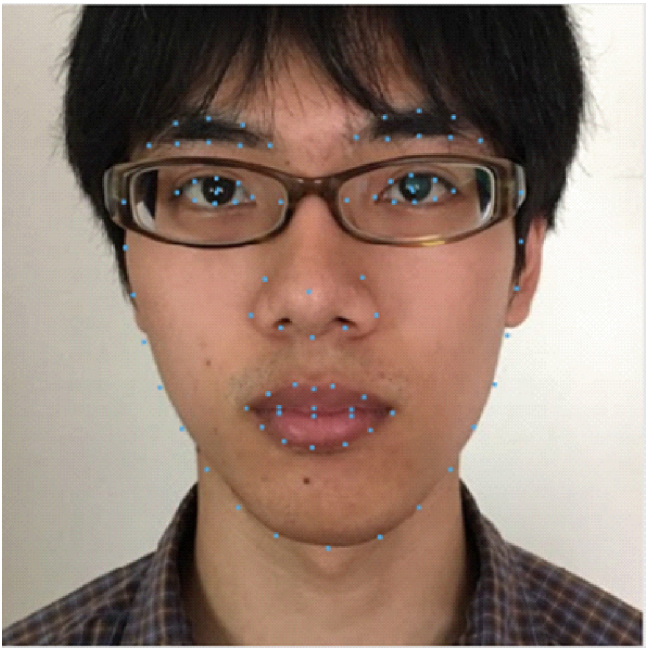

Fig. 2. Facial landmarks derived by Face++

Face recognition and analysis are the major technologies within the cognitive computing. Recently, several companies have published Web API for face recognition. For instance, Microsoft provides Face API and Emotion API [6] within the Azure Cloud Services.

Most face recognition technologies (e.g., [2,10,13]) use machine learning to detect characteristic points on a face such as eyes, nose, mouth and so on. Such points of the face are called facial landmarks. When a new picture is given to the machine, it tries to extract the facial landmarks within the picture. Figure 2 shows that facial landmarks which are extracted from my picture using the API of Face++ [4] are overlaid on the picture. Face++ API is developed based on [2]. It can derive 83 landmarks as coordinates on a picture. Moving position of these extracted landmarks to stretch and transform the picture, we can perform operations to move any specific parts of the face in the picture.

2.3 Related Work

Hinds reported the impression of the robots in case when people do group work with human-like robot or machine-like robot [3]. Hinds mentioned the more the robot of partner is human-like, people become to accept achievement and put confidence in the robot. The evaluation of Hinds's experiment was the impression on physical robots, which is different from the VA, the focus of this study. However, supposed it is the same in the point of the personification of robots,

it is considered that how personified the appearance of robots is influencing the contents of communication with robots. The VA of Fig. 1 is an animation character modelled after human, but by the VA getting more human-like appearance, there is some possibility of improving the communication between the VA and human.

3 The Proposed System

We referred to the problem of current PCC support system in Sect. 2.1. Accordingly, in this paper, we propose the system that is aimed to solve the problem, which is not conducted efficient cares for PWD by the faces of VA because the users cannot choose the VA based on his/her preferences in the existing PCC support system.

3.1 Key Idea

As a key idea to achieve more interesting communication care for a PWD, we aim to easily generate a VA based on the preferences of the individual PWD. Using the face recognition and analysis technologies, we generate a 3D facial model from a picture of the person whom the user wants to display as the VA. We then integrate the generated model to the PCC system, and create a situation where the PWD talks with the person whom the PWD want to talk with. As a result, we can implement more quality PCC, compared to the conventional artificial VA.

The existing system which is replaced by proposed system is constituted with application software which the VA is implemented in and Web API to control the application of VA. In the following sub-sections, we describe the functional requirements of the proposed system, and design of the application of VA and the service API. We also illustrate how the proposed system is integrated with the PCC support system.

3.2 Functional Requirements

Based on the analysis of the previous system, we identified that the proposed system must implement following three key features.

Feature F1 (generate a VA):
For a given picture of a face, the system can dynamically generate a VA with the face.

Feature F2 (instruct the VA to speak):
For a given text, the system can instruct the generated VA to speak the text.

Feature F3 (instruct the VA to change expression):
For a given command, the system can change the facial expression of the generated VA.

F1 implements a fundamental requirement of the system to create a VA of a familiar person of PWD. With this feature, every PWD can generate his/her own VA.

F2 implements an ability of speaking in the VA. With this feature, the VA is able to be a conversation partner.

F3 implements an ability of emotion in the VA. With this feature, the VA can behave like a human being, and produce friendly conversation with non-verbal communication.

3.3 Design of Service API and Application

Considering replacement of the previous VA system, we intend to integrate the VA with IoT and smart home. Hence, we deploy Web APIs to control the VA application in which implements the features mentioned in Sect. 3.2 as a Web service. By doing this, external applications within the PCC system can easily use the features **F1**, **F2**, **F3** via Web API. Here we define three Web APIs that define **F1**, **F2** and **F3**, respectively, and explain how to work in the VA application.

createVA(faceImageFile)
: Generates a VA from an image file which is specified by *faceImageFile*. For a given picture of a face, the system extracts landmark points of the face using a face recognition algorithm, and generates a 3D facial model. By default, we suppose that the VA does natural motions such as blink, swaying its face, and so on.

speak(type, text)
: Instructs the generated VA to speak (read) a text sentence specified by *text* with a type of voice specified by *type*. The type of synthesized voice is a parameter that is specified based on generated VA's gender and user's language. We suppose, within the API, that voice data (wav, mp3, etc.) should be synthesized from the text using a text-to-speech technology. As the VA plays back the voice data, we also suppose that the VA should synchronizes the lip motions, by stretching lips within the 3D facial model.

changeExpression (expression)
: Changes the facial expression of the generated VA by a label specified in *expression*. Based on the extracted facial landmark, the system stretches and transforms some portions of the 3D model, which change facial expressions. Typical expressions include normal, smile, angry, sad, surprised, and fear.

3.4 Integrating into PCC Support System

We suppose to deploy the above Web APIs on a Web server called *VAManager*. Figure 3 shows a sequence diagram showing how *VAManager* works within the PCC support system. Here we assume a scenario that a user creates a VA and greets to it.

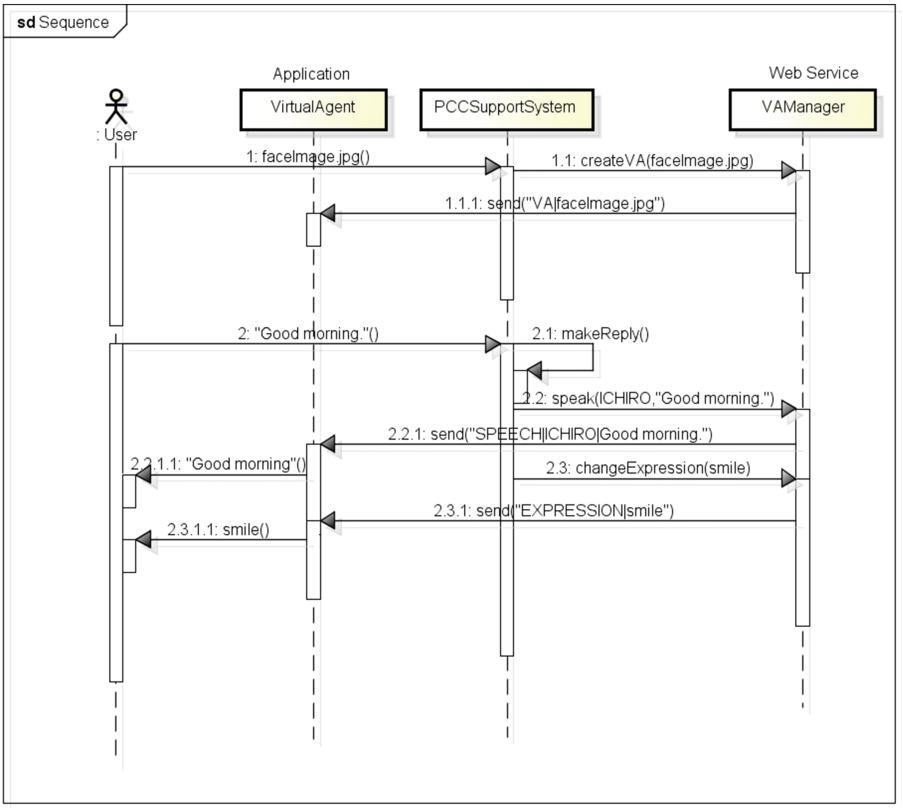

Fig. 3. The sequence of integrated system.

The user first registers an image file ("*faceImage.jpg*") to the PCC support system. The PCC support system then executes *createVA(faceImage.jpg)* to generate the VA. The VA looking like the person in *faceImage.jpg* is displayed on user's PC.

Suppose that the user says to the VA "Good morning". The PCC system recognizes the user's voice by speech recognition, generates the reply "*Good morning*", and executes *speak(ICHIRO,* "*Good morning.*"*)* to instruct the VA to say, "*Good morning*" with the synthesized voice type, "*ICHIRO*". At the same time, the system executes *changeExpression(smile)* to instruct the VA to smile. As a result, the VA on the PC screen replies to the user, "Good morning" with smile.

4 Implementation of Prototype

Based on the proposed system, we have implemented a prototype system. Technologies used in the implementation are as follows.

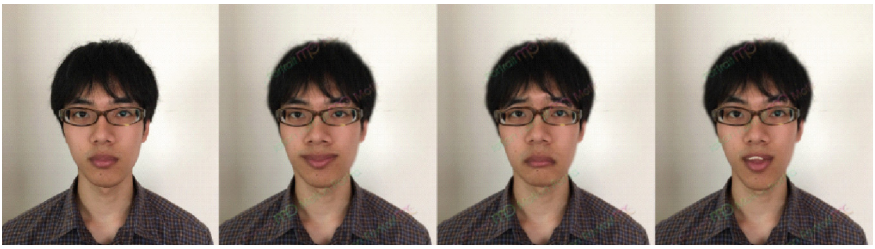

Fig. 4. Images of a virtual agent produced by the prototype system.

VAManager
- Development Language: Java
- Web Server: Apache Tomcat
- Web Service Frameworks: Apache Axis2

VirtualAgent(MPAgent)
- Development Language: C#
- MotionPortrait SDK [8]
- Bing Speech API [5]

VAManager is a Web API that is implemented by Java, and deployed on a Tomcat Web server. Based on requests from the PCC support system, it controls Virtual Agent application.

MPAgent is an application software that is implemented by C# using MotionPortrait SDK. This SDK provides powerful libraries for generating and operating 3D facial model. With the SDK, we were able to implement features **F1** to **F3** very efficiently in C#. We also use Bing Speech API within Microsoft Azure Service to synthesize voice for VA. You can get a synthesized voice with this API by specifying text and voice type defined gender and language. With this API, the system can synthesize different voice based on VA's gender and user's language. The application receives commands *VAManager* and then executes **F1** to **F3**.

Figure 4 shows example facial images in the prototype system. The images show, from left to right, the original face picture, the VA with "happy" expression, the VA with "sad" expression, the VA that is speaking.

Figure 5 shows a scene of the demonstration video talking with the VA generated with the picture of *Albert Einstein*.

The generated movie file can be watched in https://youtu.be/sXaSQJpXojc.

5 Experiment

The purpose of this experiment is to evaluate whether the faces of VA have an influence on receptivity to the care by PCC support system.

Fig. 5. A scene of the demonstration video.

5.1 Outline of Experiment

At nursing facility, we conducted evaluation experiments with the proposed system, which targets the five elderly from 74 to 99 years old, including person with dementia and person needed care/support.

In the experiment, first, we execute care program for subjects by VCG using character VA which is shown in Fig. 1, then we generate the VA by the proposed system from a picture of familiar person for each subject based on answer in advance questionnaire, "Who is your favorite singer?". After the subjects experience a short conversation with VA about three minutes, we interview subjects and get answer orally.

5.2 Result of Experiment

We got the answers to the interview from four out of five subjects. The other one was not interested in this system, so we could not get answers from her. Questions of the interview and answers for that are as follows. The same answers from different subjects are omitted.

1. What do you feel the impression of the VA generated from a picture (ex. locution, expression, accord with between voice and face)?

  - Locution and expression are not on my mind, but accord with between the voice and face of VA is a little.
  - I feel strange about it, but locution is not unnatural and accord with between the voice and face of VA is not on my mind.
  - I feel surprised and interested. Locution and expression are normal.
2. The system can create a VA if there is even a picture, whom do you want to talk with?
  - There is no one whom I talk with especially.
  - No one occurs to my mind now.
3. Which is it to talk with existing VA in Fig. 1 or real VA generated from a picture?
  - I prefer to the generated with a picture. I can also speak with the previous VA, but I want to use the new VA in the sense that I can see my family or spouse.
  - I prefer to the generated VA with a picture.
  - In the past, there is not being with faces like the previous VA. So, because the previous VA is a strange being, I prefer to the generated VA with a picture which has a human-like and familiar face.

5.3 Discussion

The discussion about each particle of the interview are as follows.

Impression of the VA
Because a voice of VA is generated by Bing Speech API, it is different from a natural voice of person in picture. There is also some possibility that some people feel uncomfortable in intonation and timing of speaking because of the synthesized voice. Compared with the previous VA based on character, it is considered that how to feel a face and expression of the VA differs by individual. In this point, according to result of the experiment, there are both opinions to concern about accord with face and voice of the VA and not to concern. Therefore, we need to improve the system by using more comfortable voice.

Person who want to display as the VA
The subjects did not mention who they want to display as VA. However, from the opinion that the user can see his family or spouse, according to the person, it is considered that the better effects can be gotten by displaying the person who is closer to user as a VA.

Easiness to talk with the VA
In comparison previous VA and proposed VA, the experiment shows the VA which has a human-like appearance by generating with picture is more acceptable than the unfamiliar character VA. All of the subject who answered the interview preferred to the VA generated with picture. However, it is considered that there is a generational problem about this from the answer of subjects. From the answer that he had not ever seen the face of the previous VA, we might be able to get different opinions from a relatively young

generation that have some opportunity to see 3D the character model as the computer spreads. That is to say, when the current young generation become to the elderly in the future, how acceptance for the appearance of the VA is different from that by the current elderly, as a result, it is considered that strangeness of the VA generated with picture might be emphasized.

According to the above, there is the case to feel uncomfortable about accord with face and voice of the VA. However, in the point of system, because the method to use a picture is highly effective, this method is expected to be useful in the scene of the care.

6 Conclusion

In this paper, we propose a system that aims more efficient person-centered dementia care with the virtual agent (VA) technology. Using the face recognition technology, the proposed system generates a 3D facial model from a given picture, integrates the model as a personalized VA of the PCC support system. Since every user can easily generate his/her own VA with a preferred face, the system can be expected to achieve more effective PCC. Then, we implemented the prototype system based on the proposed system, and experimented about effect on the face of the VA.

Acknowledgements. This research was partially supported by the Japan Ministry of Education, Science, Sports, and Culture [Grant-in-Aid for Scientific Research (B) (16H02908, 15H02701), Grant-in-Aid for Scientific Research (A) (17H00731), Challenging Exploratory Research (15K12020)], and Tateishi Science and Technology Foundation (C) (No. 2177004).

References

1. Cabinet Office, Government of Japan: Annual Report on the Aging Society, June 2017. http://www.cao.go.jp/
2. Fan, H., Cao, Z., Jiang, Y., Yin, Q., Doudou, C.: Learning Deep Face Representation. CoRR abs/1403.2802 (2014). http://arxiv.org/abs/1403.2802
3. Hinds, P.J., Roberts, T.L., Jones, H.: Whose job is it anyway? A study of human-robot interaction in a collaborative task. Hum. Comput. Interact. **19**(1), 151–181 (2004)
4. Megvii: Face++. https://www.faceplusplus.com
5. Microsoft: Bing Speech API. https://azure.microsoft.com/ja-jp/services/cognitive-services/speech/
6. Microsoft: Emotion API. https://azure.microsoft.com/en-us/services/cognitive-services/emotion/
7. MMDAgent Project Team: MMDAgent - Toolkit for Building Voice Interaction Systems. http://www.mmdagent.jp
8. MotionPortrait Inc.: MotionPortrait. https://www.motionportrait.com

9. Sakakibara, S., Saiki, S., Nakamura, M., Yasuda, K.: Generating personalized dialogue towards daily counseling system for home dementia care. In: Duffy, V.G. (ed.) DHM 2017. LNCS, vol. 10287, pp. 161–172. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58466-9_16
10. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701–1708, June 2014
11. Tamamizu, K., Sakakibara, S., Saiki, S., Nakamura, M., Yasuda, K.: Capturing activities of daily living for elderly at home based on environment change and speech dialog. In: Duffy, V.G. (ed.) DHM 2017. LNCS, vol. 10287, pp. 183–194. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58466-9_18
12. Tokunaga, S., Tamamizu, K., Saiki, S., Nakamura, M., Yasuda, K.: VirtualCare-Giver: personalized smart elderly care. Int. J. Softw. Innov. (IJSI) **5**(1), 30–43 (2016). https://doi.org/10.4018/IJSI.2017010103. http://www.igi-global.com/journals/abstract-announcement/158780
13. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 94–108. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_7