Intel researchers develop an eye contact correction system for video chats

June 26 2019, by Ingrid Fadelli

Eye contact correction: in the input frame (left) the user is looking at the screen. In the output frame (right) his gaze is corrected to look into the camera. Credit: Isikdogan, Gerasimow & Michael.

When participating in a video call or conference, it is often hard to maintain direct eye contact with other participants, as this requires looking into the camera rather than at the screen. Although most people use video calling services on a regular basis, so far, there has been no widespread solution to this problem.

A team of researchers at Intel has recently developed an eye contact correction model that could help to overcome this nuisance by restoring eye contact in live video chats irrespective of where a device's camera and display are situated. Unlike previously proposed approaches, this model automatically centers a person's gaze without the need for inputs specifying the redirection angle or the camera/display/user geometry.

"The main objective of our project is to improve the quality of video conferencing experiences by making it easier to maintain eye contact," Leo Isikdogan, one of the researchers who carried out the study, told TechXplore. "It is hard to maintain eye contact during a video call because it is not natural to look into the camera during a call. People look at the other person's image on their display, or sometimes they even look at their own preview image, but not into the camera. With this new eye contact correction feature, users will be able to have a natural face-to-face conversation."

The key goal of the study carried out by Isikdogan and his colleagues was to create a natural video chat experience. To achieve this, they only wanted their eye contact correction feature to work when a user is engaged in the conversation, rather than when they naturally take their eyes off the screen (e.g. when looking at papers or manipulating objects in their surroundings).

"Eye contact correction and gaze redirection in general, are not new research ideas," Isikdogan said. "Many researchers have proposed models to manipulate where people are looking at in images. However, some of these require special hardware setups, others need additional information from the user, such as towards what direction and by how much the redirection needs to be, and others use computationally expensive processes that are feasible only for processing pre-recorded videos."

The new system developed by Isikdogan and his colleagues uses a deep convolutional neural network (CNN) to redirect a person's gaze by warping and tuning eyes in its input frames. Essentially, the CNN processes a monocular image and produces a vector field and brightness map to correct a user's gaze.

In contrast with previously proposed approaches, their system can run in real-time, out of the box and without requiring any input from users or dedicated hardware. Moreover, the corrector works on a variety of devices with different display sizes and camera positions.

"Our eye contact corrector uses a set of control mechanisms that prevent abrupt changes and ensure that the eye contact corrector avoids doing any unnatural correction that would otherwise be creepy," Isikdogan said. "For example, the correction is smoothly disabled when the user blinks or looks somewhere far away."

The researchers trained their model in a bi-directional way on a large dataset of synthetically-generated, photorealistic and labeled images. They then evaluated its effectiveness and how users perceived it in a series of blind tests.

"Our blind testing showed that most people don't know when we turn our algorithm on or off, they see no artifacts but just feel like they have eye contact with the person they are communicating with," Gilad Michael, another researcher involved in the study, told TechXplore.

Interestingly, the researchers observed that their model had also learned to predict the input gaze (i.e., where it thought a user was looking before his/her gaze was corrected), even if was never trained to do that. They believe that this capability might be a byproduct of the model's continuous redirection of a user's gaze to the center, without specifying where a user was looking in the first place.

"The model simply inferred the input gaze so that it can move it to the center," Isikdogan explained. "Therefore, we can arguably consider the eye contact correction problem as a partial super-set of gaze prediction."

The findings gathered by the researchers also highlight the value of using photorealistic synthetic data to train algorithms. In fact, their model achieved remarkable results even if during training it relied almost entirely on computer-generated images. The researchers are far from the first to experiment with synthetic training data, yet their study is a further confirmation of its potential for the creation of highly-performing applications.

"We also confirmed that it is a good practice to keep mapping-reversibility in mind when building models that manipulate their inputs," Isikdogan added. "For example, if the model is moving some pixels from bottom left to center, we should be able to ask the model to move those back to bottom left and get an image that looks nearly identical to the original image. This approach prevents the model from modifying images beyond repair."

In the future, the system proposed by Isikdogan, Michael and their colleague Timo Gerasimow could help to enhance video conferencing experiences, bringing them even closer to in person interactions. The researchers are now planning to finalize their system so that it can be applied to existing video conferencing services.

"We put a lot of effort making sure our solution is practical and ready to be used in real products," Michael said. "We might now try to improve some of the byproduct findings of the algorithm such as gaze detection and engagement rating to enable adjacent use-cases."

More information: Leo F. Isikdogan et al. Eye Contact Correction using Deep Neural Networks, *2020 IEEE Winter Conference on*

Leo F. Isikdogan, et al. Eye contact correction using deep neural networks. arXiv:1906.05378 [cs.CV]. arxiv.org/abs/1906.05378