Video: Introduction to Face Recognition

[00:01] [slide 1] Now that you've learned some ways that we can derive information about the three-dimensional structure of the environment from visual images, and how we can analyze movement in the scene, we'll spend the rest of the semester exploring the problem of recognition, which is one of the most important ways that we use visual information, and a major focus of research in human and computer vision. In this first video, we'll briefly touch on why recognition is hard in general, and then quickly jump to the problem of recognizing faces. Here we'll consider the importance and challenges of face recognition, how good are human and computer vision systems at this task, and what are the main steps in the process of analyzing faces. The second video introduces some of the early and significant approaches to the computer recognition of faces.

[00:59] Why is object recognition difficult, in general? These photos provide a hint about a major challenge, which is that a particular category of object such as a dog or cat, trees or shoes, can appear in so many different ways in the image. It's hard to characterize the essence of each object category in a way that captures all the ways an instance of an object category might appear, while enabling us to make subtle distinctions, for example, between a dog and a fox.

[01:35] [slide 2] We're going to start our journey through recognition with a problem that you might think is an easier task because of less variability in appearance, and this is the recognition of faces. How do we detect regions of the image that might be faces, and how do we recognize the identity of people in the image? We'll consider perceptual studies that provide insights into how faces are analyzed by the human visual system, and studies from neuroscience that reveal areas of the brain that are specialized for the task of face recognition. After learning about some early methods for computer face recognition, we'll make our way to approaches based on deep learning, including those that were proposed as models of the neural processing of faces, and state-of-the-art face recognition systems built by companies like Google and Facebook. We'll see ways that these different disciplines have come together, as well as ways they've diverged. A first question we can ask is, why is face analysis important? You might pause the video for a moment and first reflect on your own experience - what kind of information can we infer from a face that's important for human social behavior?

[03:02] [slide 3] Being able to remember people we've seen before, like family, friends, co-workers, acquaintances, and famous people, is obviously important in our daily lives. We also infer important information such as gender, race, age, or family resemblance. It's a critical element of social communication - facial expressions enable us to assess a person's emotions or overall mood, and we read a lot into faces, such as a person's intentions, trustworthiness, competence or intelligence, and although subjective, we judge a person's attractiveness. Finally, the analysis of faces helps us to understand what's going on in a scene, for example, we can determine the direction of gaze from the faces of people in the photo on the left, and infer that their attention is all focused on this computer. So face analysis is an important aspect of social intelligence.

[04:10] [slide 4] Why is face recognition hard? A particular individual can appear very different, depending on factors like our point of view, different illumination in the scene, the presence of accessories like glasses, and these days - masks. The appearance of our face changes with expression, and we age, from young and adventurous to old and grumpy, and our hair changes. And most of the time, people aren't looking directly at you, they're going about their daily business, appearing with different poses, illumination, and size in the image, and in a scene that's cluttered with objects, a face may be partially hidden from view. Recognition in these unconstrained environments adds to the challenge of detecting and recognizing faces.

[05:05] [slide 5] How good are we at face recognition? I'll begin with a demonstration. Imagine that you're given a stack of cards, where each card contains one of these face images, and you're asked to sort the cards into piles that correspond to different identities, with all images of the same person in one pile and different people in different piles. We can't mimic this actual setup of the experiment here, but we can hopefully convey the basic idea. Pause the video for a moment, and scan through the photos informally, and estimate how many different people are depicted here. <pause> You may think 3 or 4, or maybe 6 or 7, or even more. The answer is two. In the actual study, the researchers used photos of two Dutch celebrities that were taken from the web. They found that subjects who were not familiar with the two celebrities, on average, created 7.5 different piles - on average, they thought there were 7 or 8 different people in the array, so don't feel bad if you were in this range too! For Dutch people who were familiar with the celebrities, the task was trivial and they were 100% correct at figuring out that there were only two people in the set of photos. The key point here is that there's an important distinction between recognition of familiar vs. unfamiliar faces. We can recognize a familiar person over a range of appearances, but if this person is unfamiliar, it's much harder to judge that two very different appearances are the same person.

[07:07] [slide 6] Psychologists assess face recognition ability in many ways. One method that has become a standard is the Cambridge Face Memory Test. This test was developed by Duchaine and Nakayama as a clinical test to diagnose prosopagnosia, which is the inability to recognize faces. Sadly, people with prosopagnosia are unable even to recognize their own family members. There are versions of this test available online if you're interested. The Cambridge Face Memory Test measures your ability to learn a new face. The images used in the study are cropped to remove hair so that the judgment is based just on the face, and the test uses three views of each individual, different lighting conditions, and noisy versions of the images. A study led by Jeremy Wilmer, a faculty member in the Psychology department at Wellesley, focused on individual differences between peoples' performance on a set of cognitive tasks that included face recognition. In addition to the Cambridge Face Memory Test, they had a test of subjects' ability to recognize famous people. The graph in the upper right corner shows data from lots of individuals. Each dot is a different subject in the experiment, and the graph shows the percentage of correct answers for the famous face recognition task vs. learning new faces using the CFMT task. And chance performance on the CFMT test is 33%, which would suggest that you can't learn new faces very well. What leaps out at you here? There's a

remarkable range of performance on both tests, from those who might casually say, “I’m not good with faces” and may be right, to people who are referred to as super-recognizers, who never forget a face. It’s important to keep in mind, when we compare performance of a computer recognition system to human performance, there’s tremendous variation in human ability to recognize faces.

[09:43] [slide 7] This also has serious implications for society, and I’d like to mention one more study that I think underscores this point. This study involves what you might think is a fairly simple task - subjects were presented a photo of a target face like the one at the top here, and an array of 10 photos like the one on the bottom that may include a different example of the target face. Subjects were asked to indicate whether the target face appears in the array, and if so, which one of the 10 photos on the bottom is the target face. Even in the best of conditions, when the faces all have the same pose and lighting, viewers were only 70% correct. As the viewing conditions are varied, performance degrades, but peoples’ base level performance is pretty bad. Think about this the next time you’re going through airport security and the agent is trying to determine whether you’re the person in your passport photo, or you’re sitting in a courtroom listening to witness testimony about a defendant they picked out of a lineup or a photo array. Society may overestimate the ability of people to recognize human faces.

[11:05] [slide 8] So how good are the best machines? First, how can we measure this? There are many publicly available databases of face images that are used as benchmarks to compare face recognition performance across systems. I mention two of them here. Labeled Faces in the Wild is a database of over 13,000 images of celebrities from the web, with almost 6,000 distinct individuals. The YouTube Faces Database has over 3,000 videos that include about 1,600 different identities. There are probably hundreds of papers comparing performance of different algorithms on databases like these. Two systems with a high impact on the field were developed by Facebook and Google. Both systems use deep learning networks that are trained to recognize faces given a vast set of training data - collections of face images that only companies like Facebook and Google can amass. These databases are private of course, and every face is labeled with the identity of the individual. A typical task used to measure performance is, given a pair of images, are they the same person or different? Studies have also evaluated human performance on these public databases, using the same task, and compared that to performance of the Facebook and Google systems. The impression one gets from the results shown here is that machines are similar or maybe better than human performance. On the right are a few examples of errors made by Google FaceNet - false accepts are cases where the machine says the adjacent pair of images are the same person when they’re not, and false rejects are actually the same person, but the machine says they’re not. We probably do better in these particular cases. Keep in mind that images in the Labeled Faces in the Wild database are not really so wild, they have limited variation in pose, illumination, and accessories like sunglasses - humans are better in these more challenging conditions. Also, if your aim is to understand how people learn to recognize faces, it’s not

through supervised training with millions of face examples - we learn new faces to a large extent from relatively little experience.

[13:57] [slide 9] What are some common applications of machine vision, of face recognition in particular? There are many applications related to security and access control, for example - verifying a person's identity to control access into a computer or secure facility, confirming identity for air travel, or bank machines. There are also applications involving visual surveillance. These are different problems that are being solved here. For problems like access control, the system is trying to perform a 1:1 matching between a user being viewed and a single individual in the database, and you can set up favorable viewing conditions for performing this match. For a task like surveillance, there's a large database of faces, and the system needs to solve a 1:N matching problem, comparing one image of a face to many individuals in the database and trying to select the best matches, and the camera may be viewing people at a distance, covertly, and under varying pose, illumination, and occlusion.

[15:18] [slide 10] Face recognition capability has become available on social media platforms, but its original introduction, for example, on Google Glass, was met with great controversy over issues of privacy and security, leading Google to temporarily ban the integration of face recognition on its platforms back in 2013. Companies are trying now to be more responsible about access to personal face images, and at least enabling users to turn off face recognition on their devices. Having a deeper understanding of things like facial expressions, and how they're generated and interpreted, is having an impact on the creation of technologies like humanoid robots that can mimic human facial expression more realistically, and has also led to the creation of more naturalistic renderings of faces using computer graphics, which is important for the entertainment industry.

[slide 11] As we examine computational methods for processing face images, remember that there are many aspects of face analysis, and we'll focus on the first two, starting with face identification and later considering how we can detect faces in the image in the first place. The other aspects here are things you might consider exploring in your final presentation.

[slide 12] On a final note, our visual system is so tuned in to detecting faces that we see them everywhere.