Faces of Facebook:

Or, How The Largest Real ID Database In The World Came To Be

Alessandro Acquisti, Ralph Gross, and Fred Stutzman

Carnegie Mellon University

BlackHat 2011 – White Paper

In Steven Spielberg's movie rendition of "Minority Report" (a short story by author Philip K. Dick), advertising technology has become so advanced by 2054 that remote retina scans can be used for personalizing electronic billboards to match the interests and backgrounds of passers-by. That future may be much closer than imagined, and perhaps even more ponderous. The research we describe in this white paper investigates how the combination of online social network data and commercially available off-the-shelf facial recognition applications can be used to successfully identify individuals online (for instance, across different sites, such as a social network and a dating site) and offline (for instance, on the street), as well as to infer -- in real-time -- additional, sensitive information about those individuals. The application of such re-identification techniques to brick-and-mortar and electronic commerce may be enthralling; at the same time, its privacy implications are unnerving.

Our research is based on the near-future, inevitable convergence of two trends: 1) the slow but steady improvements in computer facial recognition algorithms, and 2) the avalanche of personal photos that Internet users post publicly online, often in an identified format. For instance, Facebook has become the largest repository of photos on the Internet. Since Facebook has been enforcing (albeit unevenly) a verified, single identity policy (under which Facebook users are required to create profiles under their real first and last names, and the usage of pseudonyms can lead to one's account deletion), Facebook profiles may soon become the largest identity database in the world; a sort of *de facto* "Real ID" that markets and IT, rather than government and regulation, have created.

The existence of such a large and semi-openly accessible database of identities makes it plausible to consider scenarios whereby members' profile data can be used to re-identify individuals both online (for instance, on websites where their photos are uploaded without their names) and offline. We designed three experiments to test the feasibility and effectiveness of using social network profiles for individual re-identification. The first two experiments tested the possibility of identifying individuals both online and offline. The last experiment tested the possibility of inferring even more personal and sensitive information about a stranger merely by combining, in real time, facial recognition algorithms and access to online resources through a simple mobile device.

In the first experiment, we used images from Facebook profiles that were publicly accessible directly via popular search engines (such as Google), and successfully re-identified a significant proportion of pseudonymous profiles on a dating site popular in the United States.

In the second experiment, we used publicly available images from a social networking site popular among college students to identify individuals walking around the campus of a North-American academic institution. Passers-by were invited to participate in the experiment by sitting in front of a webcam for the time necessary to take three photos, and then by completing a short survey. While a participant was completing her survey, her photos were uploaded to a computing cluster and matched against a database of images from profiles on the social networking site. Thereafter, the participant was presented with the images that the facial recognizer had ranked as the most likely matches for her photograph. The participant was asked to complete the survey by indicating whether or not she recognized herself in each of the images. Using this method we re-identified a significant proportion of participants.

In the third experiment, we inferred personal information from a subject's social network profile in real time, after recognizing her face through an application installed on a common mobile phone device. We then linked to her, through her face, additional personal information found (or inferred, through data mining) online, and displayed that information on the phone. This example of an "augmented reality" application embodies both the promises and the significant perils raised by the upcoming combination of facial recognition, social networks data, cloud computing, and mobile devices.