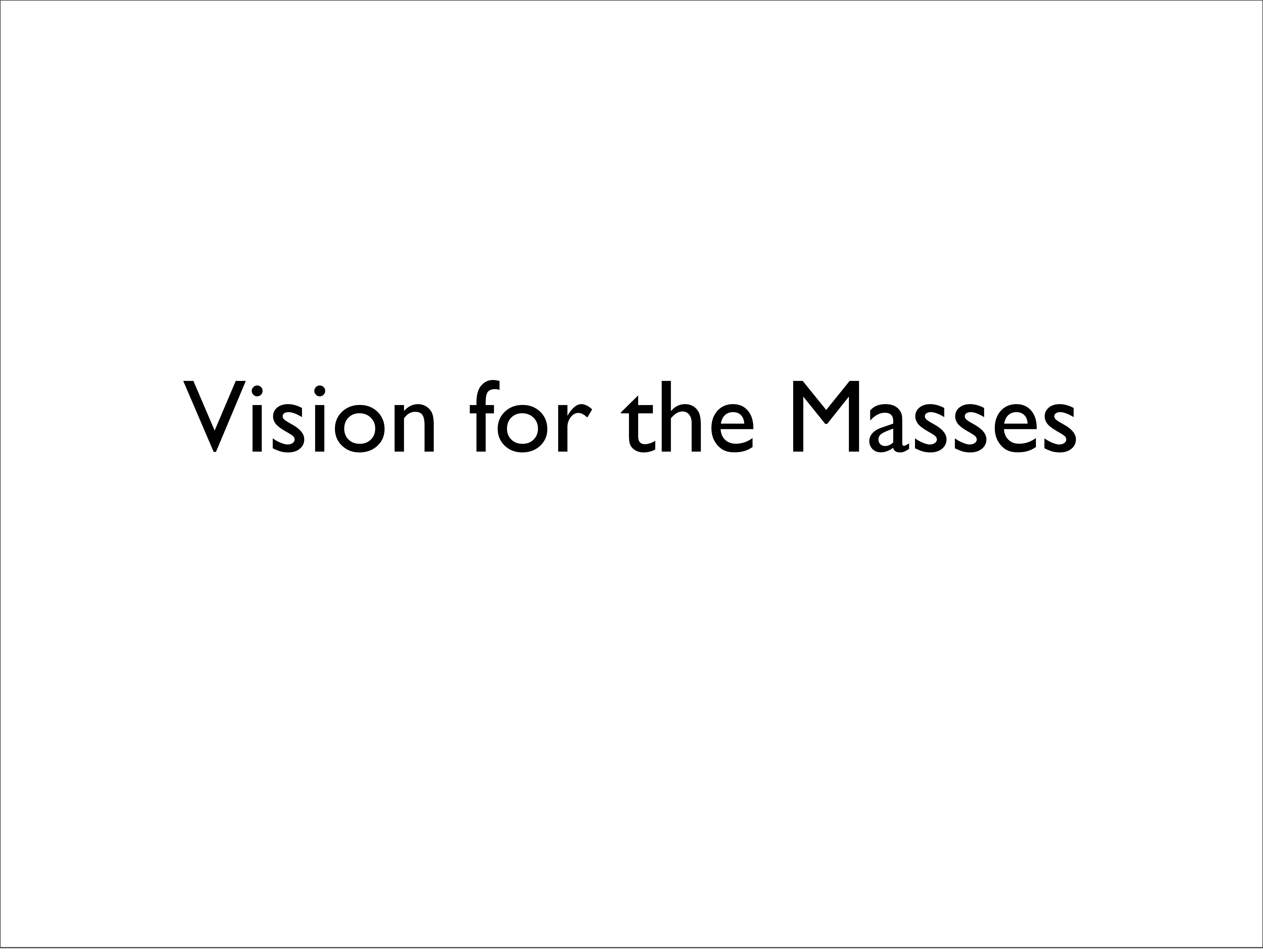
Computer Vision on Tap

Kevin Chiu, Ramesh Raskar
Camera Culture @ MIT Media Lab



Vision for the Masses

Vision on Tap is about bringing Vision to the Masses. For example, let's say my mom wants to know when a nearby parking spot becomes available.

What is she going to do?

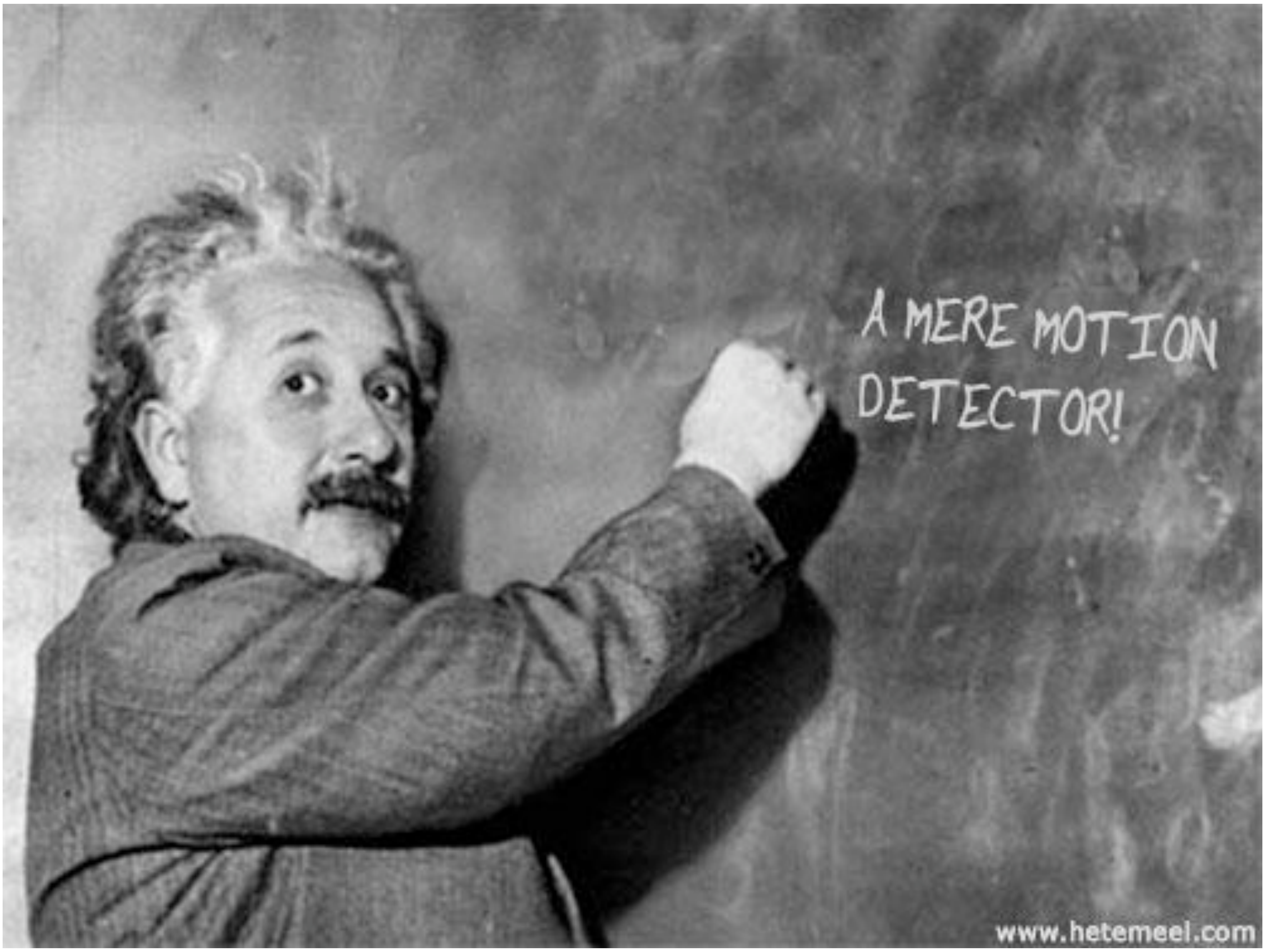
Demo

VisionOnTap.com

All she needs to do is visit a url, aim an attached camera at the parking spot, select the spot she's interested in watching, and set up remote alerts to tell her when the car moves. Then, she can go on with her with her life knowing that her computer is watching her spot for her.

About the Demo

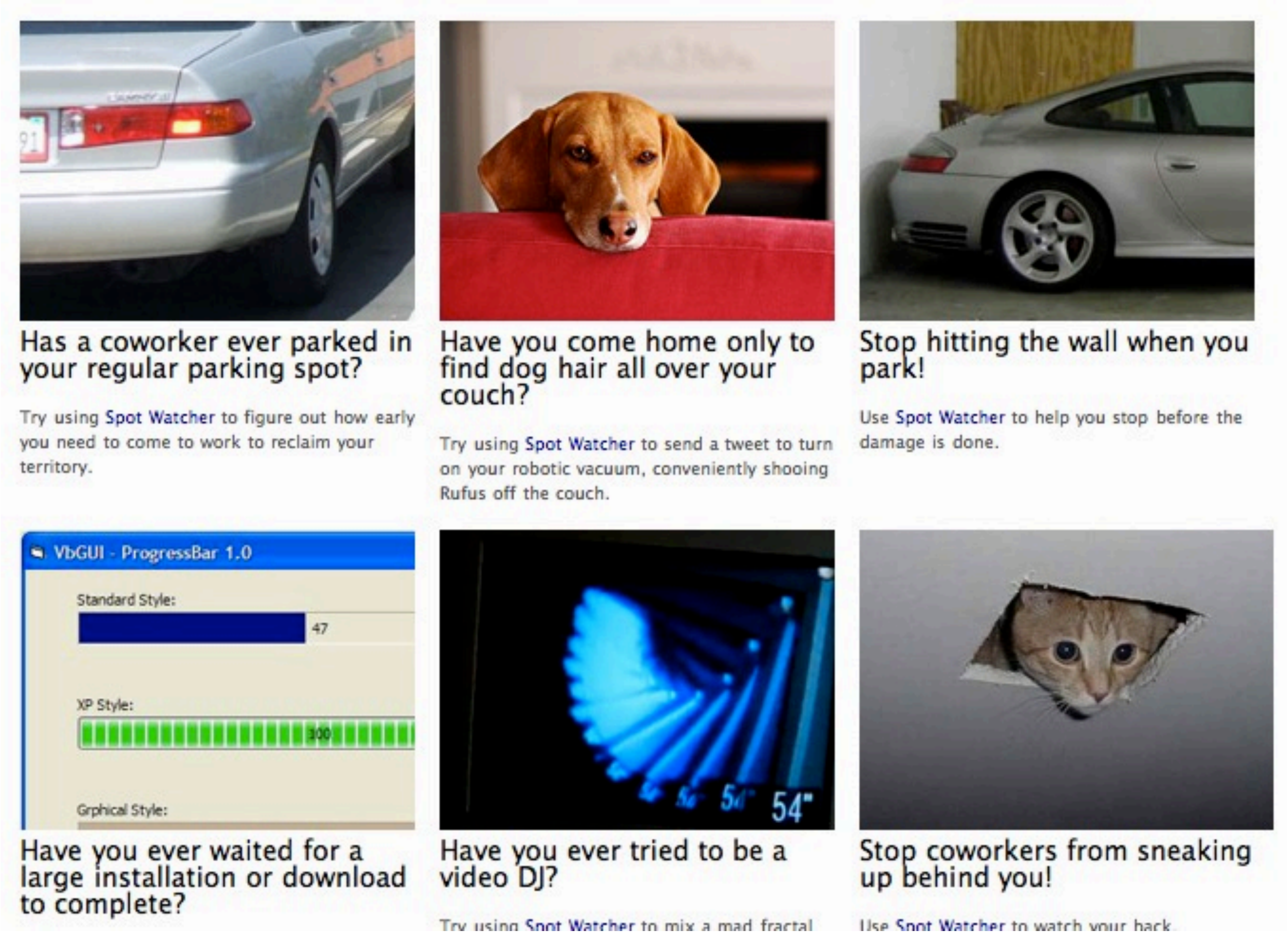
As you saw, she didn’t have to download any software, no images were shared, and her problem was solved.



For those of us familiar with computer vision, the technical portion of this demonstration may have been trivially simple.

However, the focus should be on how this technology empowers people who don't consider themselves vision experts.

Potential Applications



Has a coworker ever parked in your regular parking spot?

Try using Spot Watcher to figure out how early you need to come to work to reclaim your territory.

Have you come home only to find dog hair all over your couch?

Try using Spot Watcher to send a tweet to turn on your robotic vacuum, conveniently shooing Rufus off the couch.

Stop hitting the wall when you park!

Use Spot Watcher to help you stop before the damage is done.

Have you ever waited for a large installation or download to complete?

Have you ever tried to be a video DJ?

Try using Spot Watcher to mix a mad fractal

Stop coworkers from sneaking up behind you!

Use Spot Watcher to watch your back.

So what we saw was a very simple demo of motion detection, but even with that simple application, there are many uses.

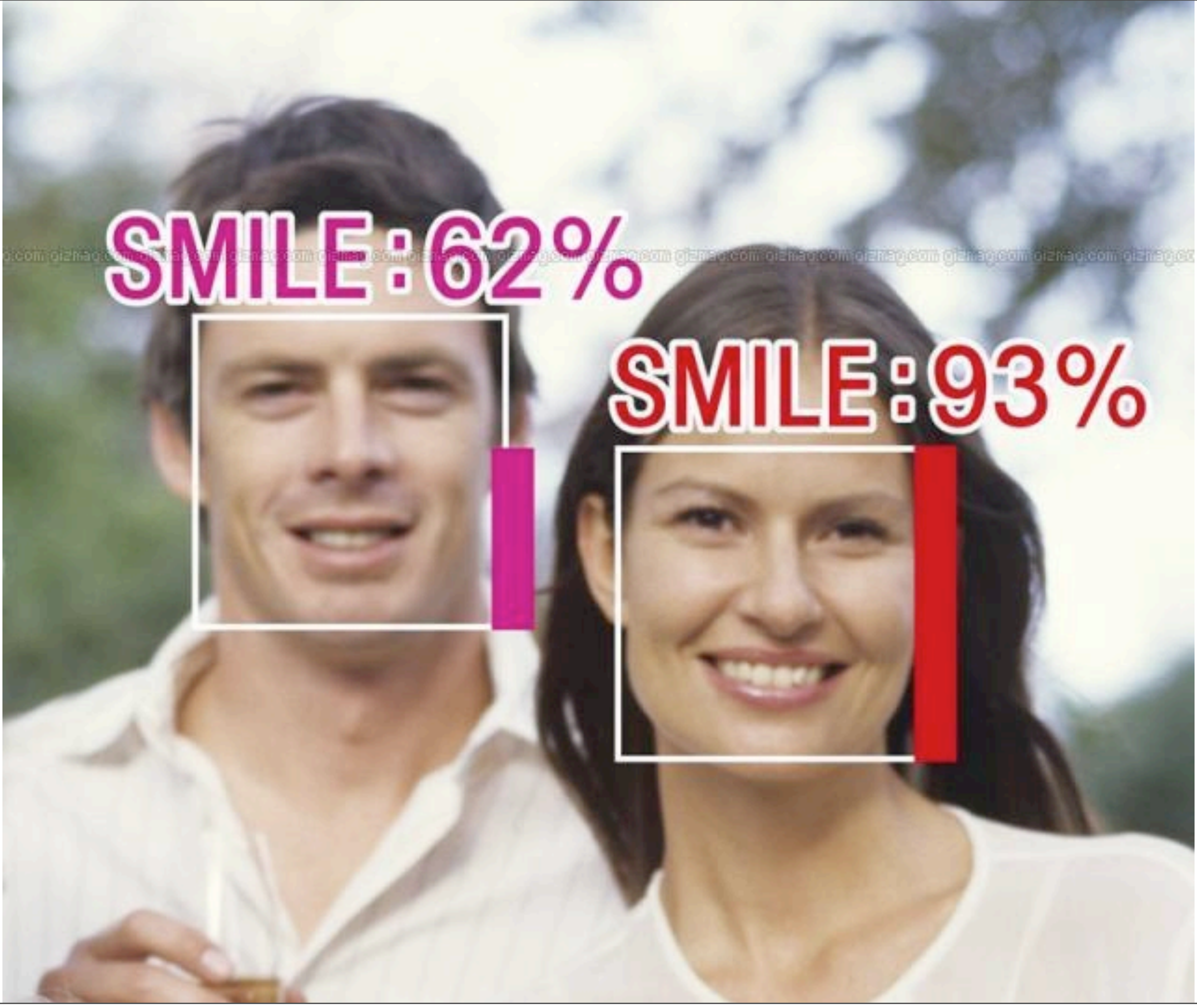
What about free range kids?

Alice, a proud new mother with a toddler, could use it as an enhanced baby monitor, keeping watch over her daughter and alerting her if her free range offspring enters an off limits zone.

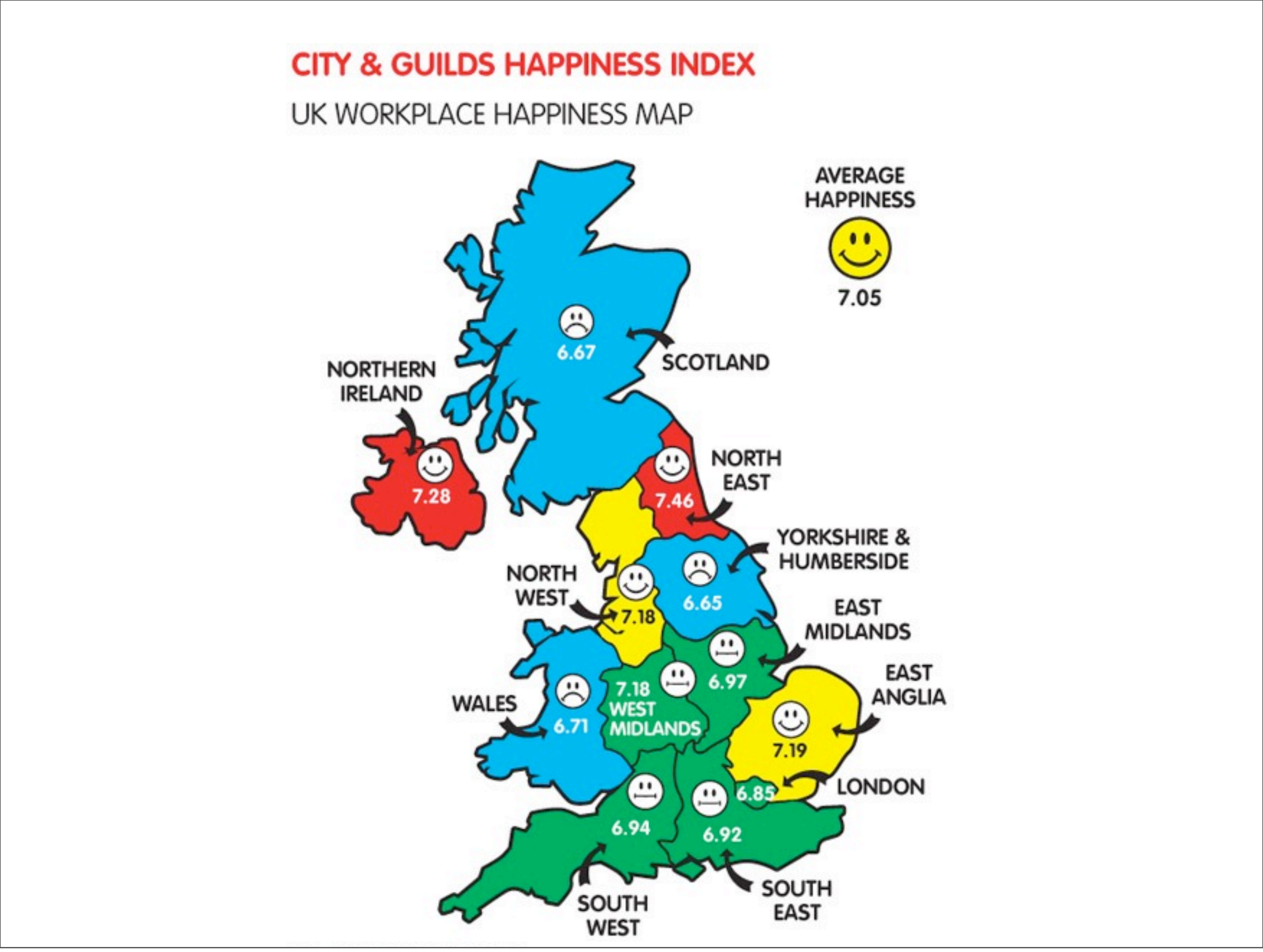
Bob, the technical tinkerer, could keep his dog off his couch by sending an alert over the web to his robotic vacuum cleaner.

Just imagine the applications consumers will come up with for more sophisticated tools...

They could catch dangerous speeders with optical flow.



smile detector

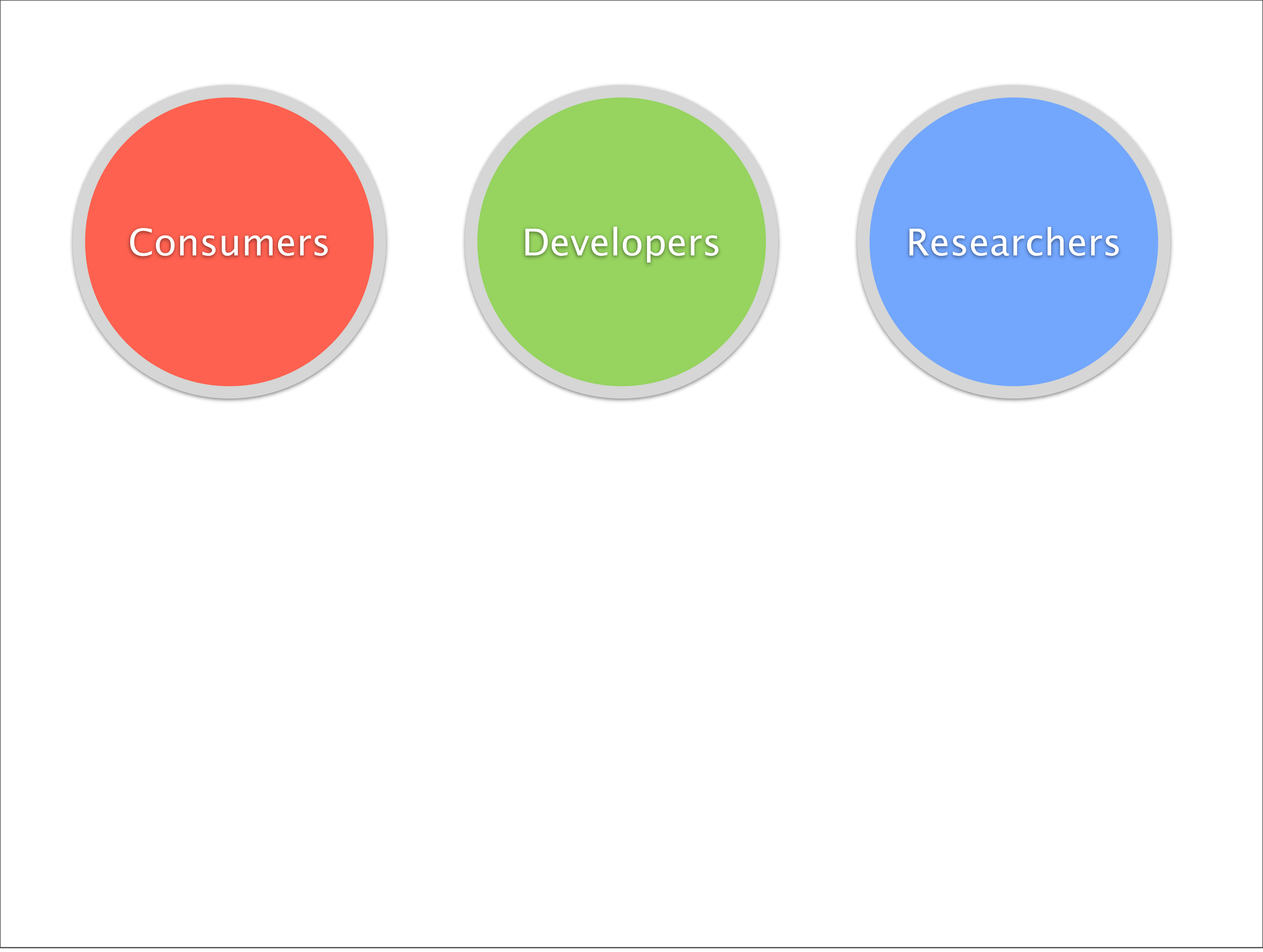


By merely reporting smile counts using real time video analysis, progressive workers can combine their results to make a happiness map of their workplace and compare them to workplaces across the country.

Keep in mind that no images need to be shared to make this work.

Accessibility

At its core, Vision on Tap is about making computer vision accessible to anyone with access to the Internet.



Consumers

Developers

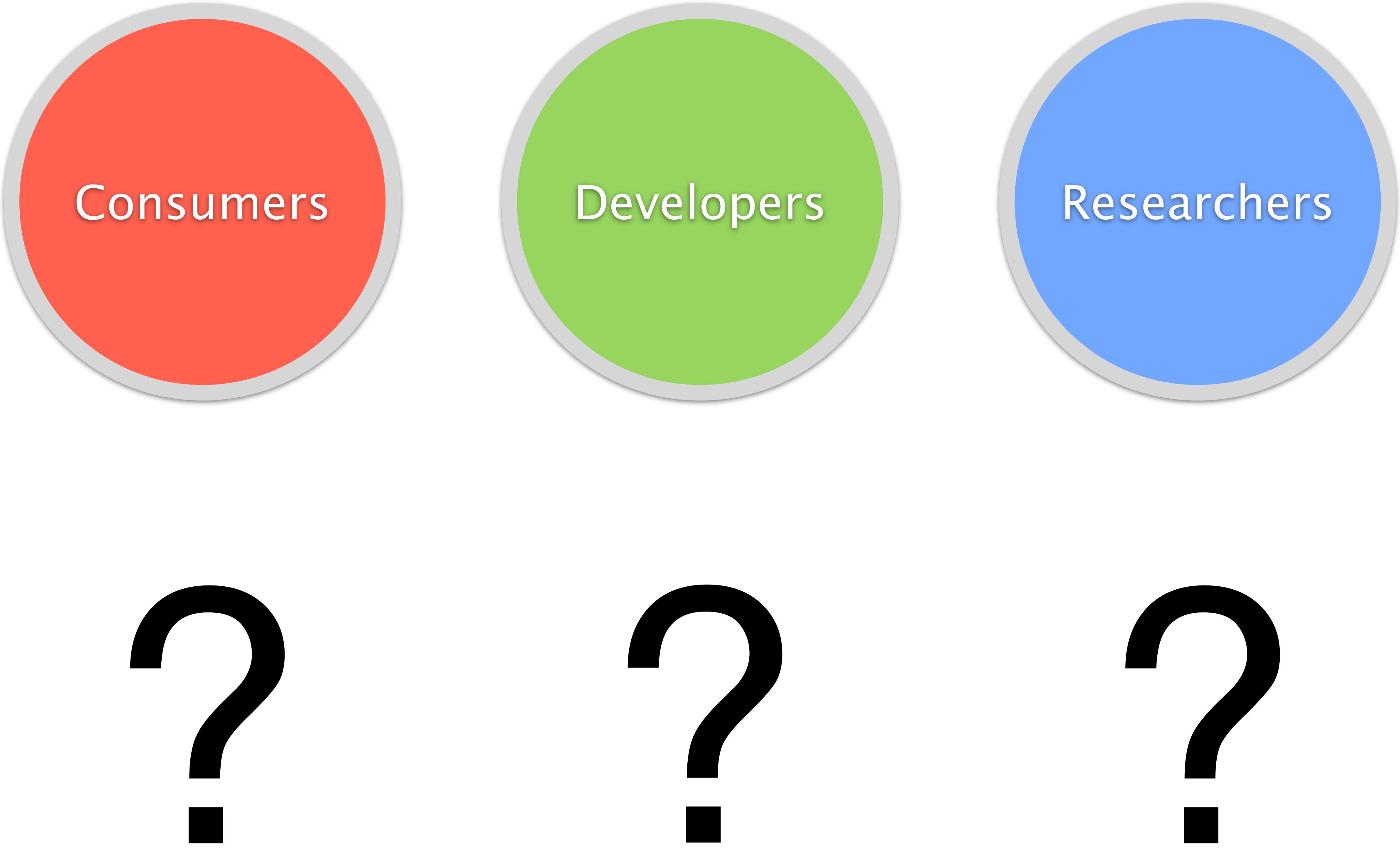
Researchers

The target audience can be thought of as three non-exclusive groups including consumers, developers, and researchers.

Consumers want an easy to use solution to their problems.

Developers want to be able to create solutions for their customers without having to learn a lot of domain specific knowledge concerning computer vision.

Researchers want to increase their exposure, impact, and opportunities for innovative collaboration.

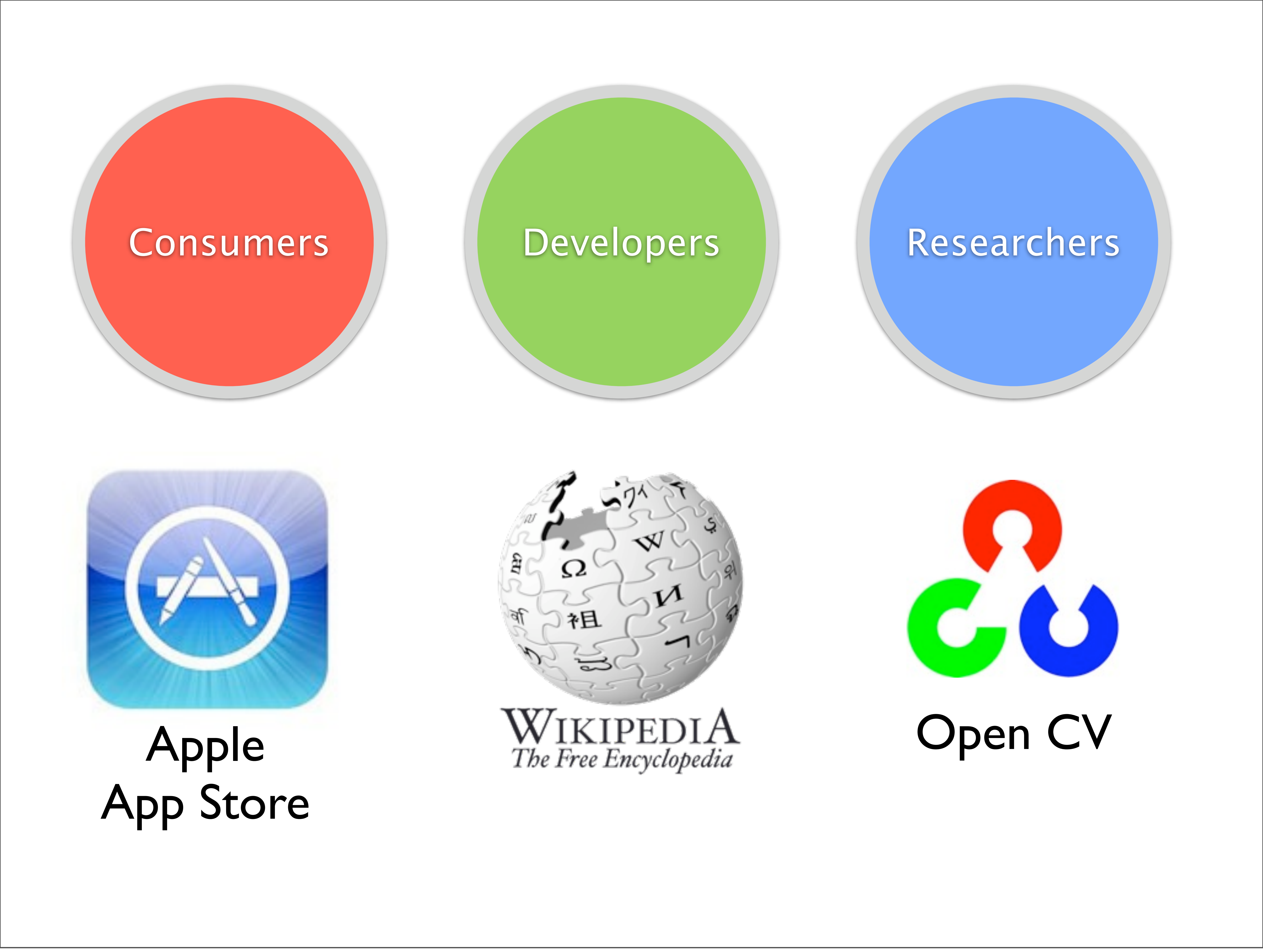


Consumers Developers Researchers

? ? ?

The problem with the way things are today is that there is no definitive place for consumers to find and customize solutions to their problems, many of which can be solved using even the most basic computer vision algorithms. Developers don’t have a platform for fast solution creation, and Researchers don't have a platform they can use to serendipitously combine their efforts.

We're all doing excellent research in computer vision, but it seems that a good portion of it stays within the vision community where normal people aren't able to use them in their daily lives.



Consumers

Developers

Researchers

Apple
App Store

WIKIPEDIA
The Free Encyclopedia

Open CV

So, how will Vision on Tap solve these problems? For consumers, it will be a vast library of new solutions. For developers, it will be a prime resource for collaboratively customizing computer vision modules, and for researchers, it'll be an open forum for collaboration.

The State of the Art

Scene Completion Using Millions of Photographs (CMU)

80 Million Tiny Images (MIT)

Photo Tourism (UW)

ImageNet (Princeton)

Utility data annotation with Amazon Mechanical Turk (UIUC)

LabelMe (MIT)

Blind Vision (Avidan et al @ MERL)

Privacy Cam & PICO (Boult et al @ UCCS)

Over the last few years, our nascent field has encountered many new challenges as it breaks new ground.

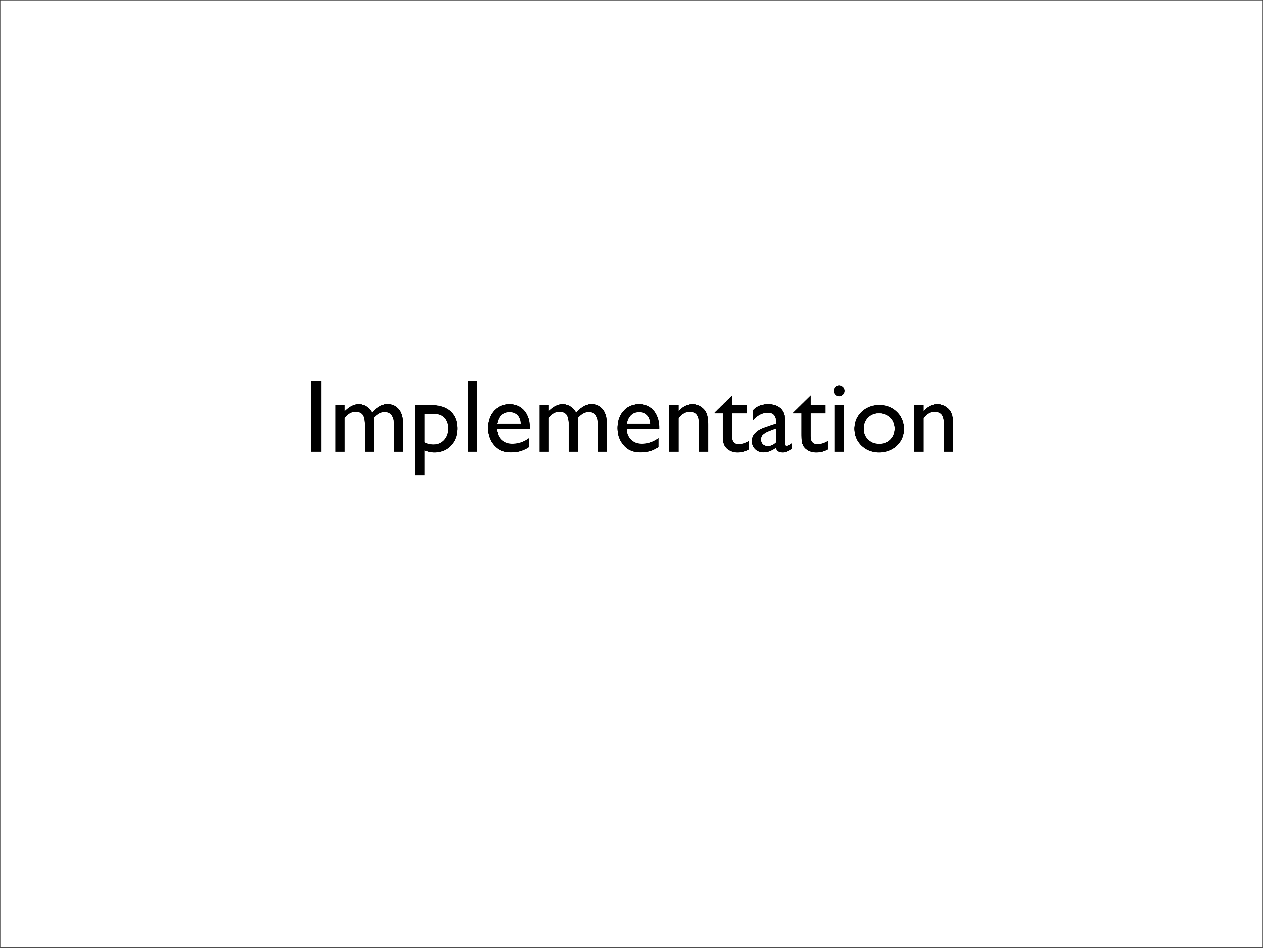
Aspects inherit to the web, including massive user-generated data repositories, loosely coupled distributed services, and dealing with privacy expectations, have mixed with our established research directions in vision to produce brand new opportunities uniquely characteristic of Internet Vision.

There are two seemingly established new directions that have made themselves apparent in Internet Vision.

One recent direction in internet vision focuses on using vast data stores on the internet to greatly enhance basic techniques. These include projects such as Scene Completion Using Millions of Photographs (CMU), 80 Million Tiny Images (MIT), Photo Tourism (UW), and ImageNet (Princeton).

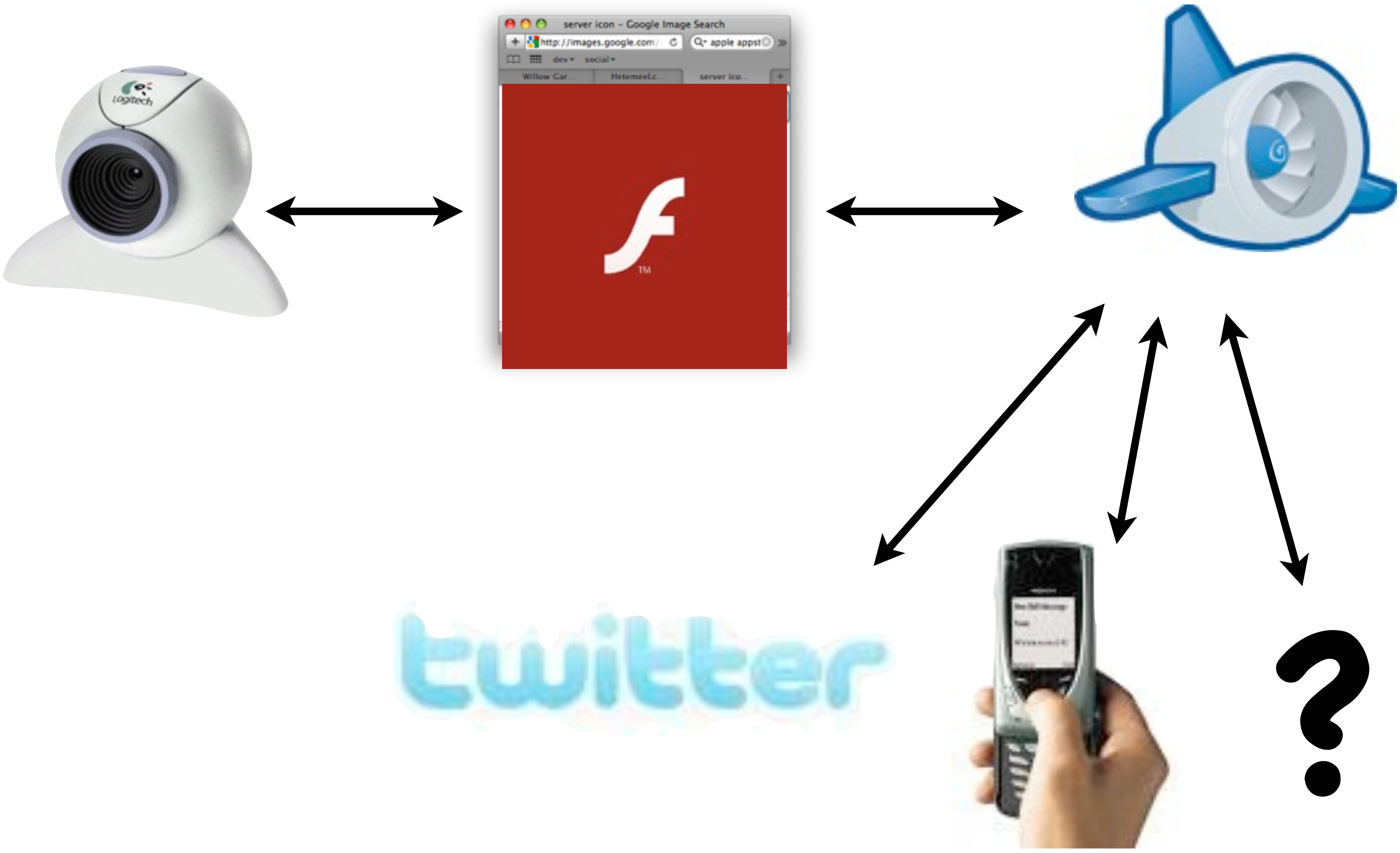
Another direction involves using distributed human effort online. These include projects such as LabelMe (MIT) and Utility data annotation with Amazon Mechanical Turk (UIUC).

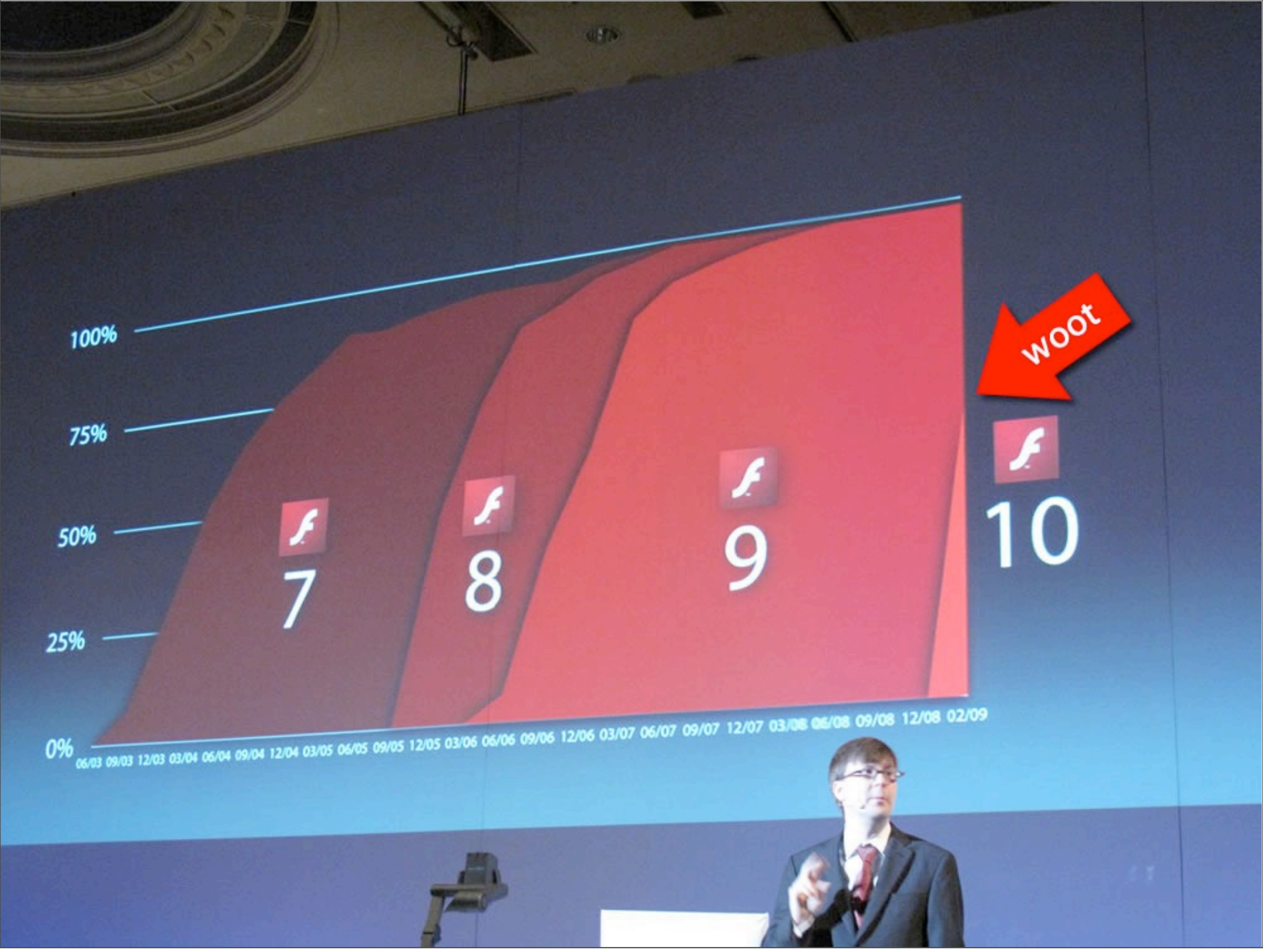
We're focusing on a nascent direction complementary to these existing efforts. Primarily, we aim to answer the question, what can you do with computer vision as a web service, and what are the challenges and opportunities involved? What are the possibilities for both consumers and creators? We're further venturing in a less explored space pioneered by Blind Vision at MERL and PICO and PrivacyCam at UCCS.



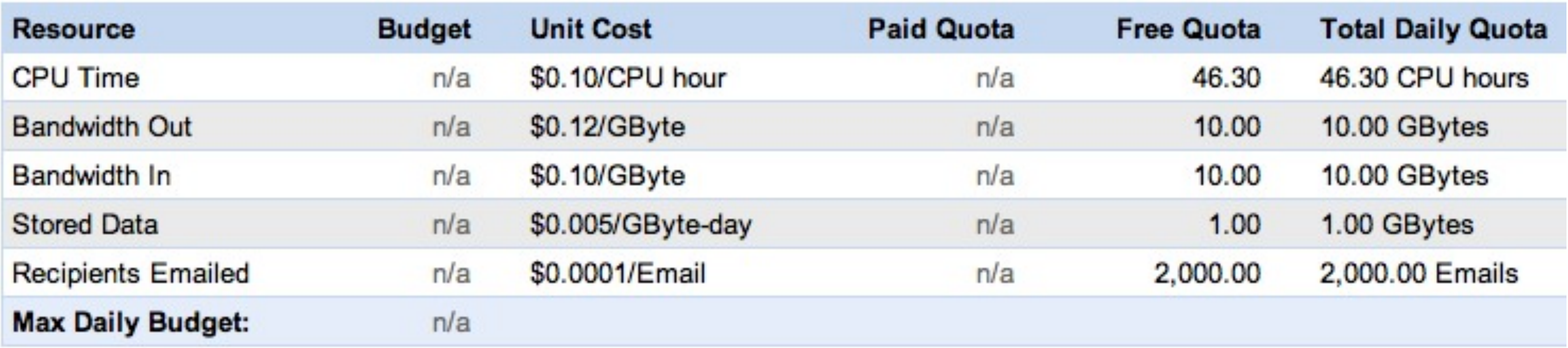
Implementation

In this section, we cover Flash and App Engine, a great way to create an extremely scalable website. We wanted users to be able to use our applications without extra installations. Flash gives us a rich platform to install on top of, and it’s already present on about 99% of clients. App Engine is a great way to create an extremely scalable website.


Logitech
server icon – Google Image Search
http://images.google.com/
apple appst
dev
social
Willow Gar...
Hetemeel.c...
server ico...
TM
twitter
?


woot
100%
75%
50%
25%
0%
7
8
9
10
06/03 09/03 12/03 03/04 06/04 09/04 12/04 03/05 06/05 09/05 12/05 03/06 06/06 09/06 12/06 03/07 06/07 09/07 12/07 03/08 06/08 09/08 12/08 02/09

Google App Engine



Resource	Budget	Unit Cost	Paid Quota	Free Quota	Total Daily Quota
CPU Time	n/a	$0.10/CPU hour	n/a	46.30	46.30 CPU hours
Bandwidth Out	n/a	$0.12/GByte	n/a	10.00	10.00 GBytes
Bandwidth In	n/a	$0.10/GByte	n/a	10.00	10.00 GBytes
Stored Data	n/a	$0.005/GByte-day	n/a	1.00	1.00 GBytes
Recipients Emailed	n/a	$0.0001/Email	n/a	2,000.00	2,000.00 Emails
Max Daily Budget:	n/a				

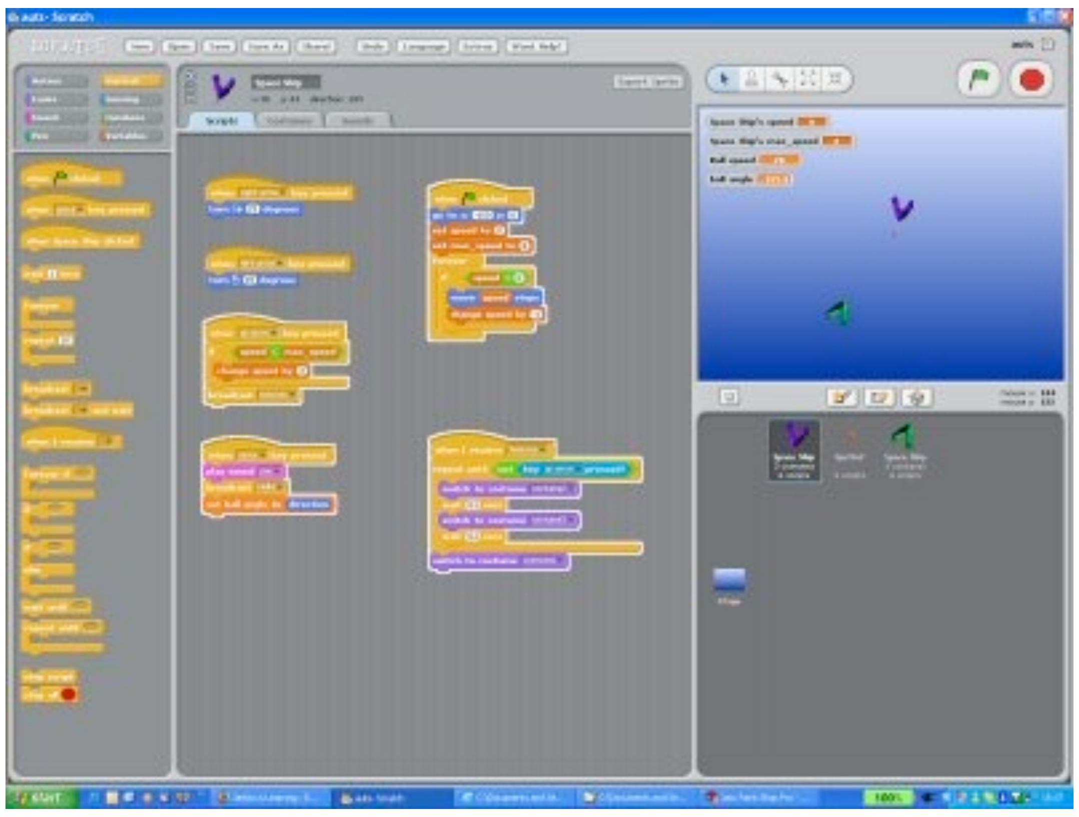
Future Directions

For consumers, developers, and researchers

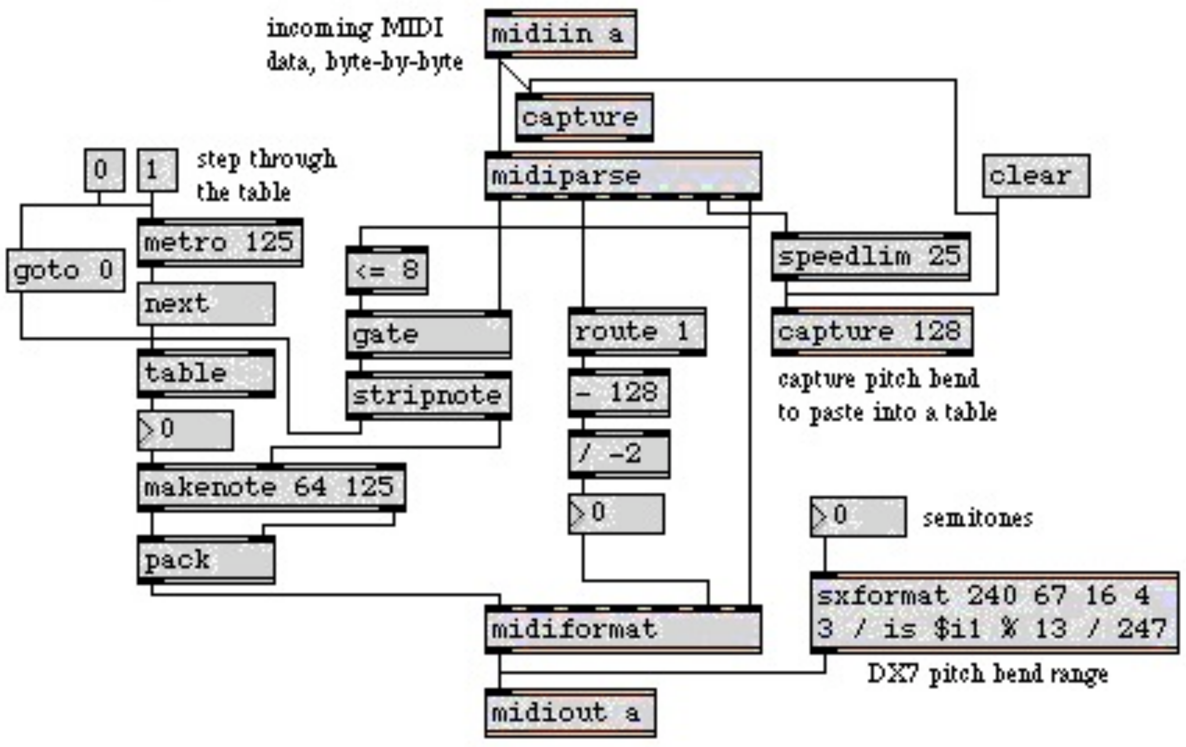
Future directions to address consumer needs will include a focus on growing the vision library and creating a full web platform where users can search for, find, and mash up solutions...

Consumers

SCRATCH



Max/MSP

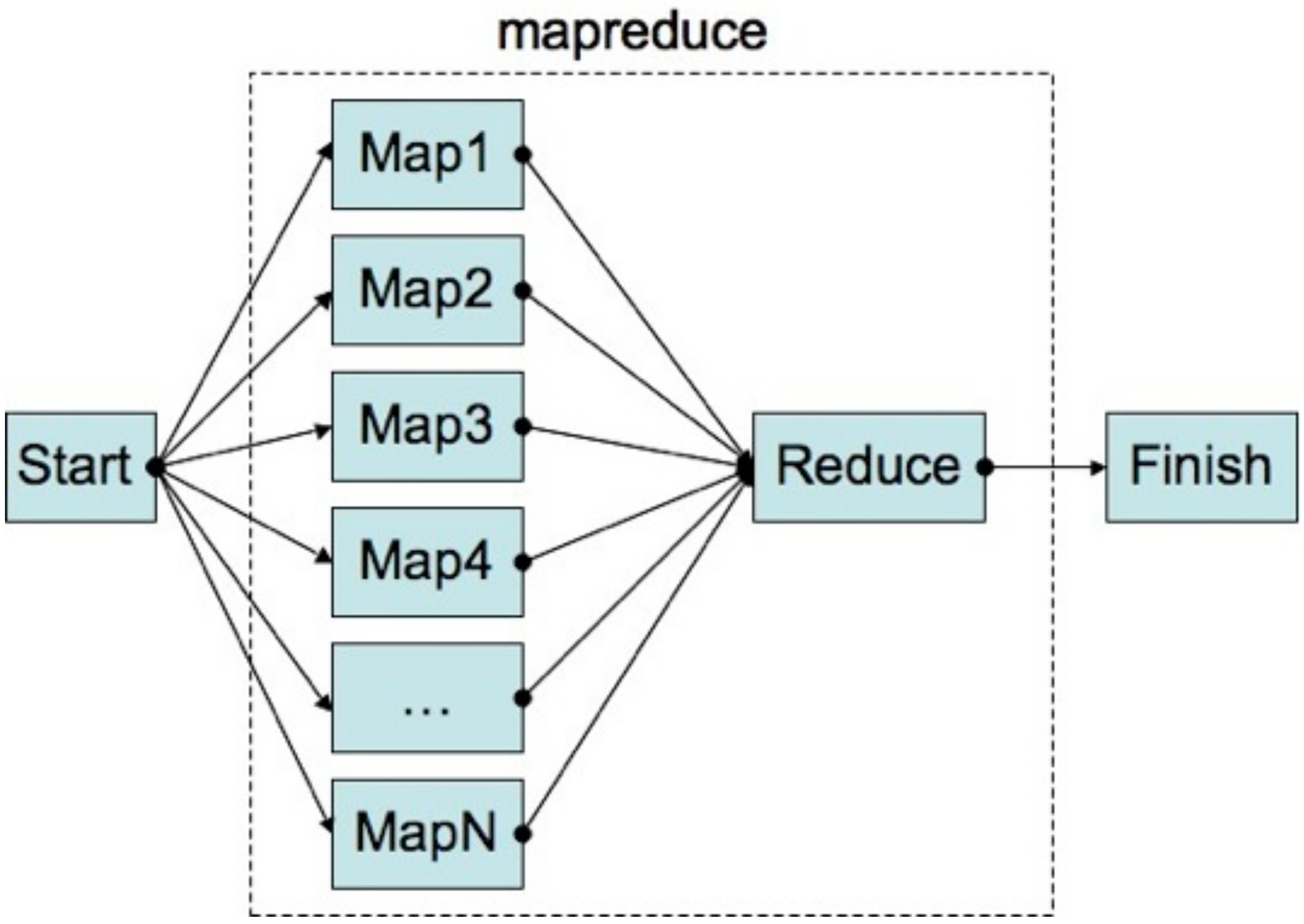


... using a simplified programming environment in the spirit of Scratch, a programming platform for kids, and Max/MSP, a dataflow oriented programming environment in which logic modules can be composed graphically.

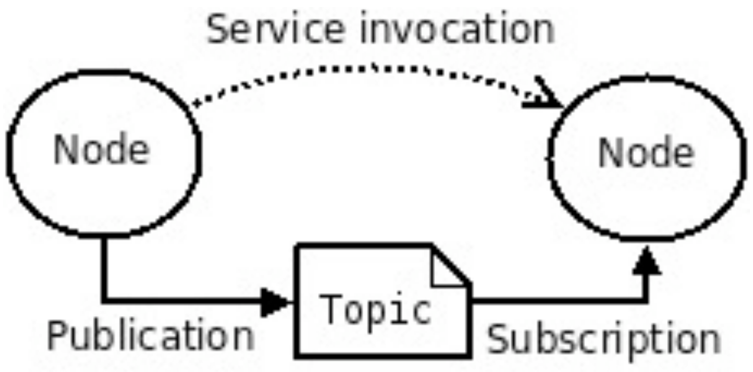
Modules

Input	Processing	Output
Local Cameras	Optical Flow	Twitter
Online Videos	Structure from Motion	Email
Remote Sensors	Smile Detection	SMS
...	...	...

Developers



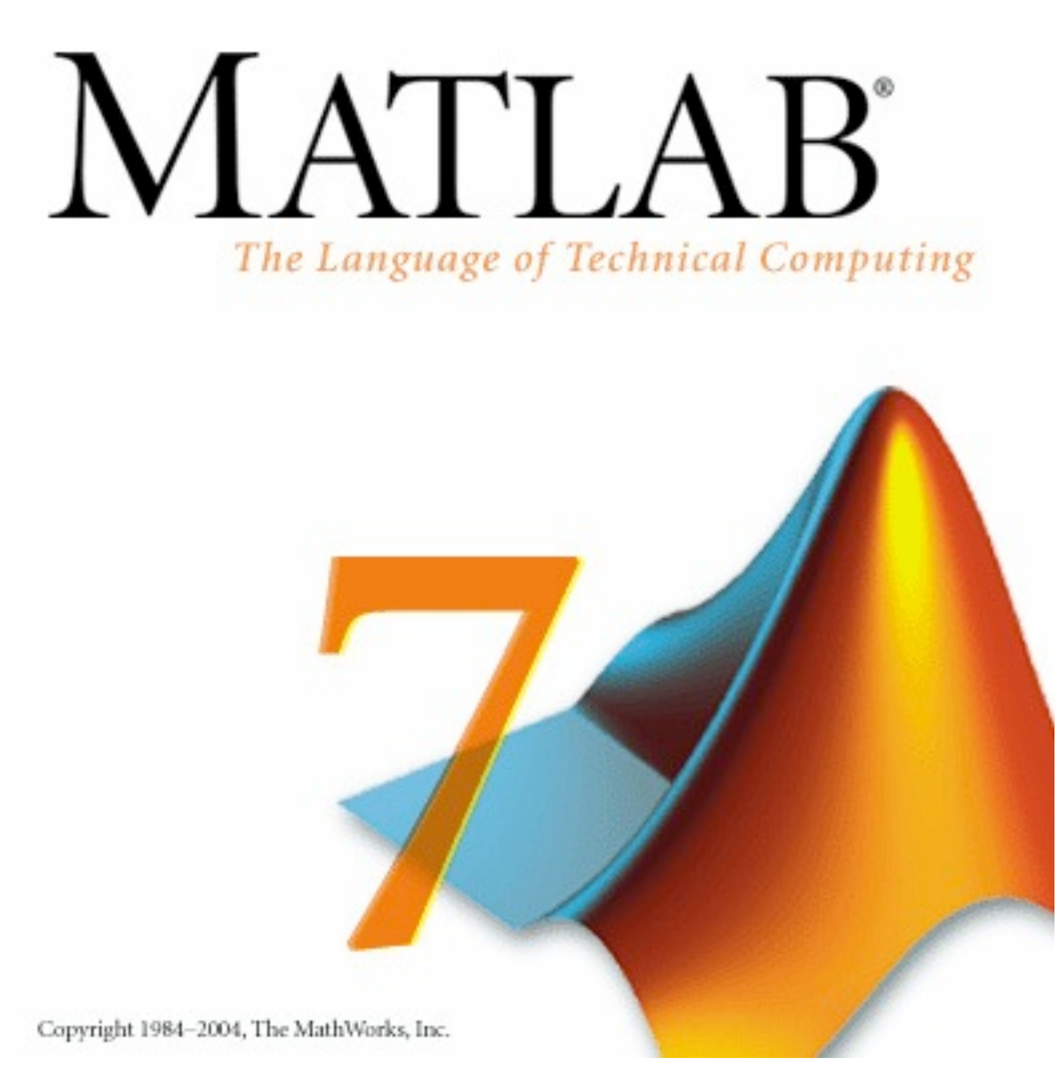
ROS



For developers, we will further explore in-browser distributed computation possibilities, for example, using flash video players to do distributed object recognition. Additionally, we will pursue the development of a ROS-like modular computation system with a Max/MSP-like graphical interface for both developers and advanced end users.

ROS is the modular robotic operating system from Willow Garage and Max/MSP is a visual data flow programming interface platform from Cycling '74.

Researchers


MATLAB®
The Language of Technical Computing
7
Copyright 1984–2004, The MathWorks, Inc.

Open CV

For researchers, we aim to provide tools to convert existing material, such as matlab or opencv-based implementations, to Flash as seamlessly as possible.

Challenges & Limitations

- **Privacy** - People do not understand that no images are shared.
- **Platform** - Adobe Flash vs Other
- **Internet** - Not always available
- **Cost** - Laptop + Camera more expensive than dedicated
- **Speed** - Interpreted vs native code
- **Adoption** - needs bootstrapping!

Many challenges await us.

Meeting privacy expectations is a challenge in cases where we rely on users video streams to do computation. Despite being clear with users that no images are being sent back to our system, users remain camera shy. Additionally, companies may be concerned that algorithms that they would like to keep proprietary might be leaked to the public. Currently, we're only focusing on user-side privacy, and rely on compilation obfuscation to protect researcher and developer privacy. In the future, we aim to construct mechanisms to respect the privacy of all users as in Blind Vision from MERL, but at the moment, we fall short of this ideal.

Our current direction aligns us with a single platform, Flash, that we do not yet know will meet our needs. Alternative environments, such as Silverlight, Android, or iPhone OS, are yet to be explored.

Our technology relies on the constant availability of the internet, which may not be the case universally.

The total cost of a flexible solution using our system might be more expensive than purchasing a dedicated solution. For example, a beam break detector for garage doors.

Speed is a challenge, but recent advances such as Adobe Alchemy and Google Native Client, which allow code to run at native or near-native speeds in the browser, may be a way to meet this challenge.

There is also a challenge related to adoption of these services. There needs to be an extensive library of applications before users will think of visionontap.com as their default destination for camera applications. Many people don't even know that computer vision might prove to be very useful for everyday life.

VisionOnTap.com

- Computer Vision as a real time web service
  - In-Browser Adobe Flash, Google AppEngine
- Consumers
  - Privacy - no images are shared
  - Ease of use - no special software
- Developers & Researchers
  - Platform for collaboration, innovation, and dissemination

Kevin Chiu

Ramesh Raskar

cameraculture.media.mit.edu

Many challenges await us.

Meeting privacy expectations is a challenge in cases where we rely on users video streams to do computation. Despite being clear with users that no images are being sent back to our system, users remain camera shy. Additionally, companies may be concerned that algorithms that they would like to keep proprietary might be leaked to the public. Currently, we're only focusing on user-side privacy, and rely on compilation obfuscation to protect researcher and developer privacy. In the future, we aim to construct mechanisms to respect the privacy of all users as in Blind Vision from MERL, but at the moment, we fall short of this ideal.

Our current direction aligns us with a single platform, Flash, that we do not yet know will meet our needs. Alternative environments, such as Silverlight, Android, or iPhone OS, are yet to be explored.

Our technology relies on the constant availability of the internet, which may not be the case universally.

The total cost of a flexible solution using our system might be more expensive than purchasing a dedicated solution. For example, a beam break detector for garage doors.

Speed is a challenge, but recent advances such as Adobe Alchemy and Google Native Client, which allow code to run at native or near-native speeds in the browser, may be a way to meet this challenge.

There is also a challenge related to adoption of these services. There needs to be an extensive library of applications before users will think of visionontap.com as their default destination for camera applications. Many people don't even know that computer vision might prove to be very useful for everyday life.