Heat of the Moment: Characterizing the Efficacy of Thermal Camera-Based Attacks

Keaton Mowery (UC San Diego)
Sarah Meiklejohn (UC San Diego)
Stefan Savage (UC San Diego)

Code-based access control

Code-based access control

Code-based access control

Code-based access control

Code-based access control

The problem: what if there is a camera watching you type in your code?

Filming keypads

The solution: just shield the keypad!

Filming keypads

The solution: just shield the keypad!

Filming keypads

The solution: just shield the keypad!

Filming keypads

The solution: just shield the keypad!

Another problem: this only protects the code while it is being typed, not after

Filming keypads

The solution: just shield the keypad!

Another problem: this only protects the code while it is being typed, not after

Turns out heat is transferred in the process of entering the code, heat residue is left after code entry

Filming keypads

The solution: just shield the keypad!

Another problem: this only protects the code while it is being typed, not after

Turns out heat is transferred in the process of entering the code, heat residue is left after code entry

Our attack: this residue can then be recorded by a thermal camera

Previous work

Previous work

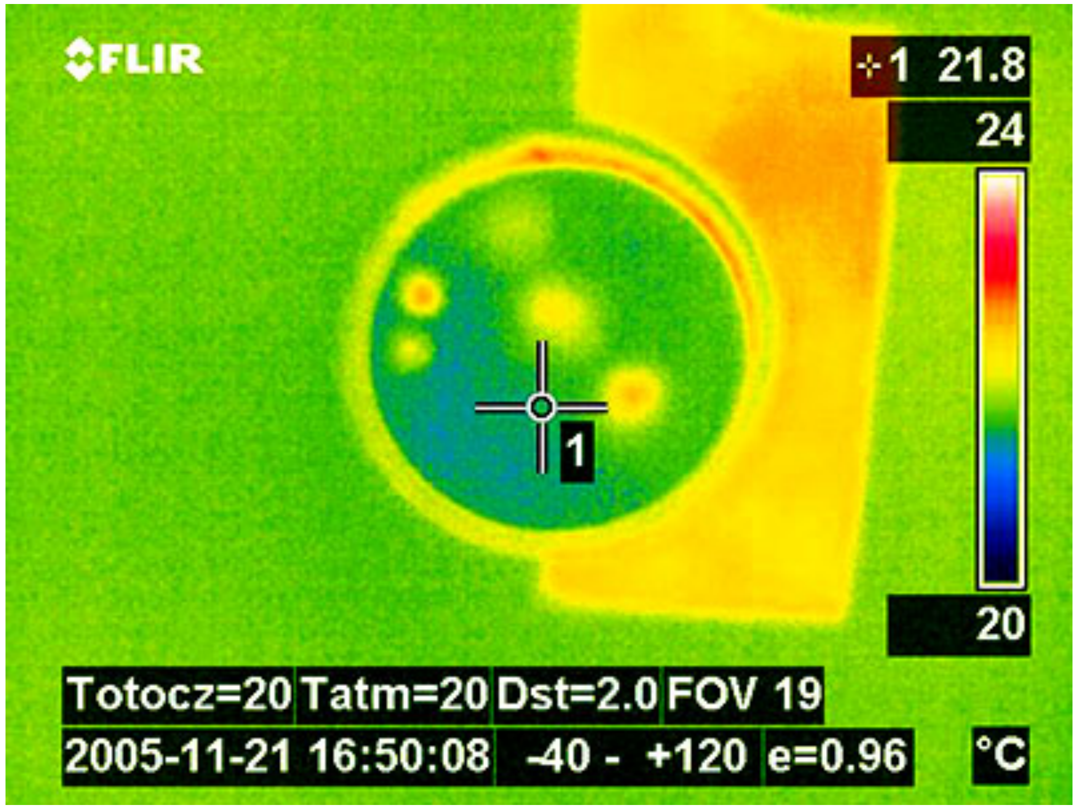
Feasibility of this attack was demonstrated in 2005 by Michał Zalewski

Previous work

Feasibility of this attack was demonstrated in 2005 by Michał Zalewski

Previous work

Feasibility of this attack was demonstrated in 2005 by Michał Zalewski

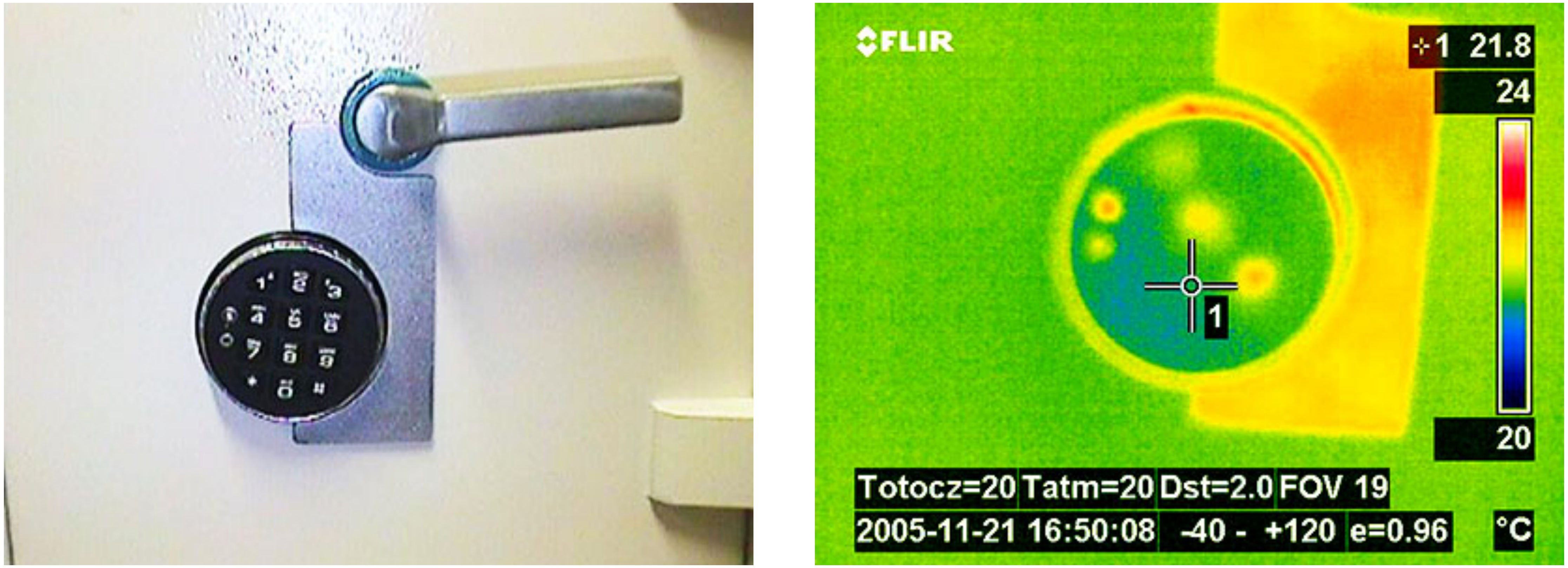


FLIR
1 21.8
24
20
Totocz=20 Tatm=20 Dst=2.0 FOV 19
2005-11-21 16:50:08 -40 - +120 e=0.96
°C
1

(images from lcamtuf.coredump.cx/tsafe)

Previous work

Feasibility of this attack was demonstrated in 2005 by Michał Zalewski


FLIR
1 21.8
24
1
20
Totocz=20 Tatm=20 Dst=2.0 FOV 19
2005-11-21 16:50:08 -40 - +120 e=0.96
°C

(images from lcamtuf.coredump.cx/tsafe)

He was able to retrieve thermal residue for between five and ten minutes after code was entered

This work

This work

We broaden the picture by considering different:

This work

We broaden the picture by considering different:

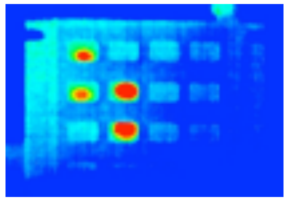
- Keypad materials (metal vs. plastic)

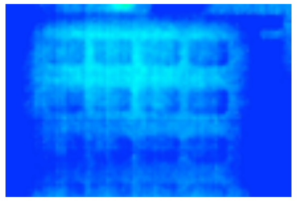
This work

We broaden the picture by considering different:

- Keypad materials (metal vs. plastic)

- Keypad users (cold- vs. warm-blooded, etc.)



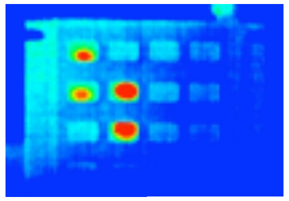


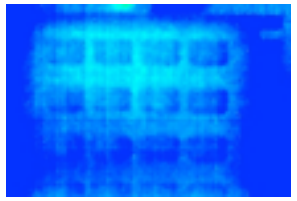
This work

We broaden the picture by considering different:

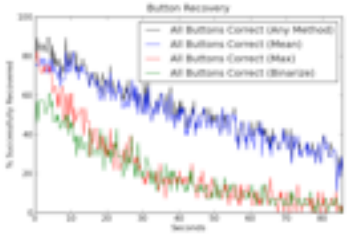
- Keypad materials (metal vs. plastic)

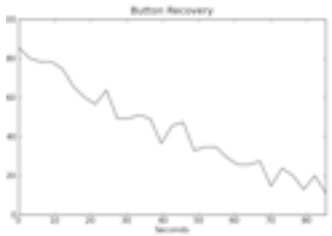
- Keypad users (cold- vs. warm-blooded, etc.)





- Review methods (automated vs. visual inspection)



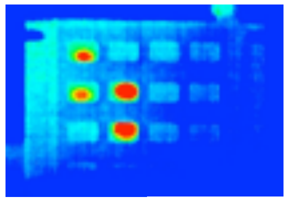
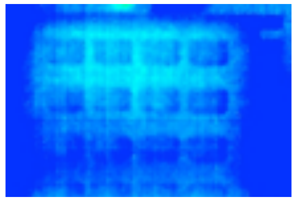


This work

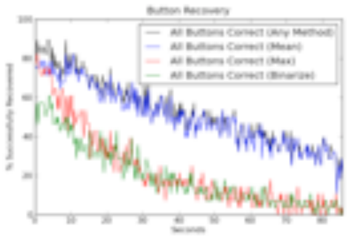
We broaden the picture by considering different:

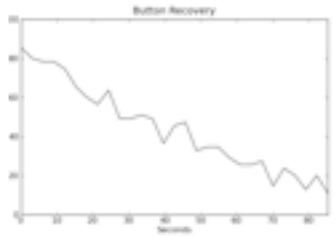
- Keypad materials (metal vs. plastic)

- Keypad users (cold- vs. warm-blooded, etc.)

- Review methods (automated vs. visual inspection)





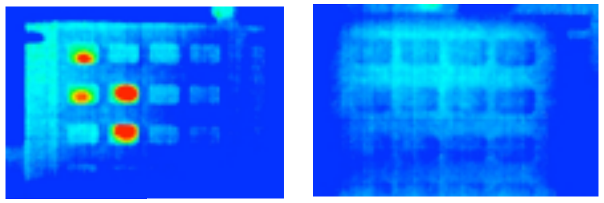
- Degrees of success (exact code vs. partial information)

This work

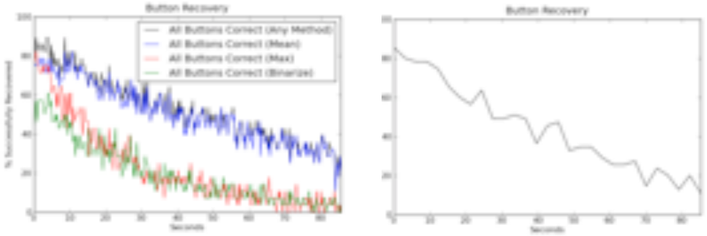
We broaden the picture by considering different:

- Keypad materials (metal vs. plastic)

- Keypad users (cold- vs. warm-blooded, etc.)



- Review methods (automated vs. visual inspection)



- Degrees of success (exact code vs. partial information)

Find that results vary substantially as we change above variables

Outline

Outline

Experiment design

Outline

Experiment design

Camera data

Outline

Experiment design

Camera data

Analyzing the data

Outline

Experiment design

Camera data

Analyzing the data

Conclusions

Outline

Experiment design

Camera data

Analyzing the data

Conclusions

Our setup: equipment

Our setup: equipment

FLIR A320 IR camera

320 x 240 resolution

$18,000 to purchase

$2,000/month to rent

Operates at 9Hz

Our setup: equipment

FLIR A320 IR camera

320 x 240 resolution

$18,000 to purchase

$2,000/month to rent

Operates at 9Hz

Metal ATM keypad

Our setup: equipment

FLIR A320 IR camera

- 320 x 240 resolution
- $18,000 to purchase
- $2,000/month to rent
- Operates at 9Hz

Metal ATM keypad

Plastic ATM keypad

Our setup: getting things ready

Our setup: getting things ready

Set keypad in a vise and camera on a tripod across from it

Our setup: getting things ready

Set keypad in a vise and camera on a tripod across from it

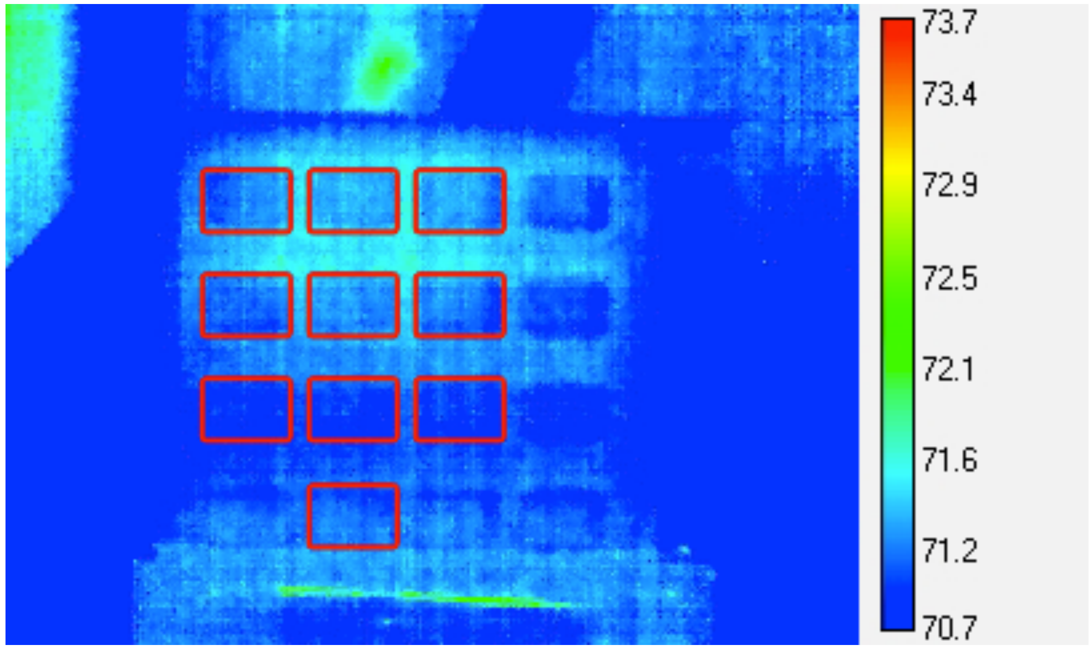
Worked at two different distances: 14 and 28 inches

Our setup: getting things ready

Set keypad in a vise and camera on a tripod across from it

Worked at two different distances: 14 and 28 inches

Used software to indicate ten regions of interest on the keypad (0-9)



Our setup: code entry

Our setup: code entry

At each distance, had 21 people type in 27 different codes

Our setup: code entry

At each distance, had 21 people type in 27 different codes

- Wanted to allow for different body temperatures, key-pressing styles, etc.
- 7 of these codes contained repeats (e.g., 6688 or 8728)

Our setup: code entry

At each distance, had 21 people type in 27 different codes

- Wanted to allow for different body temperatures, key-pressing styles, etc.
- 7 of these codes contained repeats (e.g., 6688 or 8728)

Filmed the keypad for 3 seconds before code entry, then 100 seconds after, recorded 3 frames per second

Outline

Experiment design

Camera data

Analyzing the data

Conclusions

Filming metal was a complete failure!

Filming metal was a complete failure!

Brushed metal acted as a thermal mirror, hard to even get any reading

Filming metal was a complete failure!

Brushed metal acted as a thermal mirror, hard to even get any reading

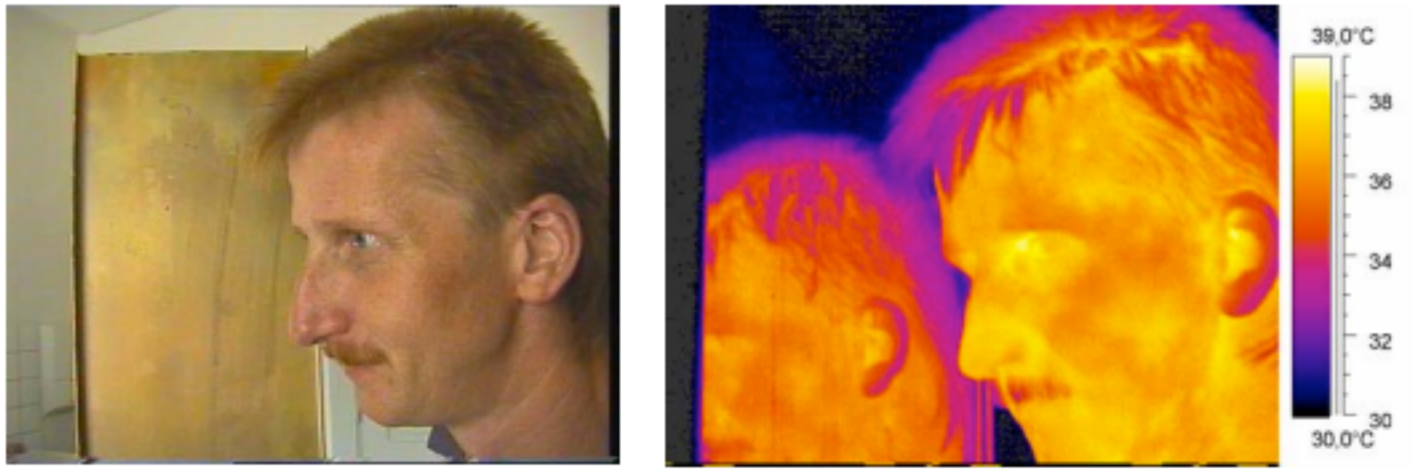


Figure 5. An oxidized old brass plate with a lot of surface roughness in the 1µm scale or below is scattering light diffusely for visible light, but at least in part specularly for thermal IR radiation of λ≈ 10µm.

(images from
“Identification and suppression of thermal reflections in infrared thermal imaging,”
Henke et. al.,
InfraMation 2004.)

Filming metal was a complete failure!

Brushed metal acted as a thermal mirror, hard to even get any reading

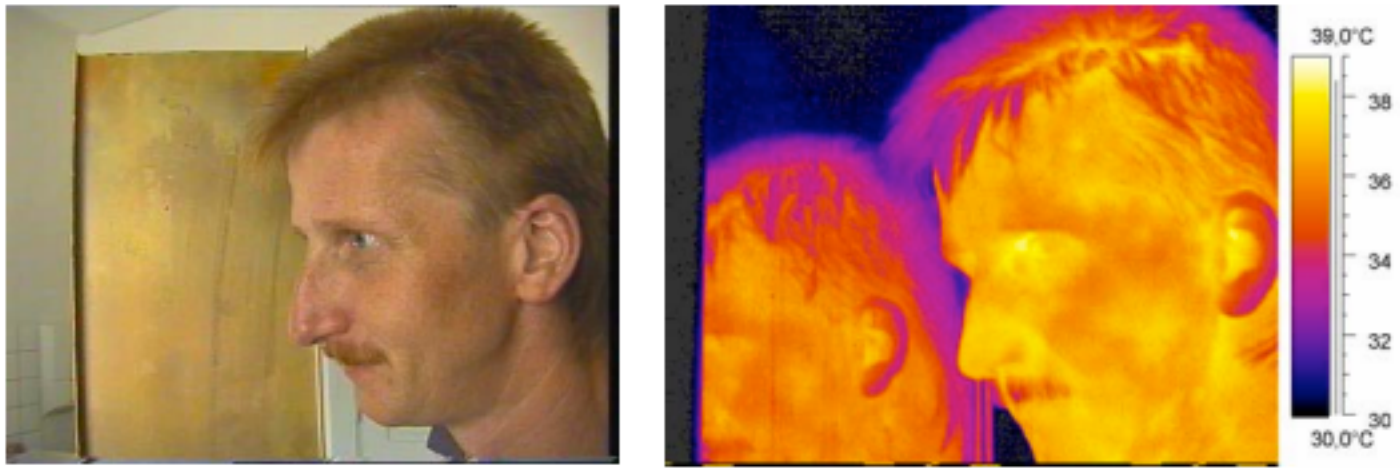


Figure 5. An oxidized old brass plate with a lot of surface roughness in the 1μm scale or below is scattering light diffusely for visible light, but at least in part specularly for thermal IR radiation of $\lambda \approx 10\mu m$.

(images from
“Identification and suppression of thermal reflections in infrared thermal imaging,”
Henke et. al.,
InfraMation 2004.)

High conductivity of metal meant residue spread within seconds

Filming metal was a complete failure!

Brushed metal acted as a thermal mirror, hard to even get any reading

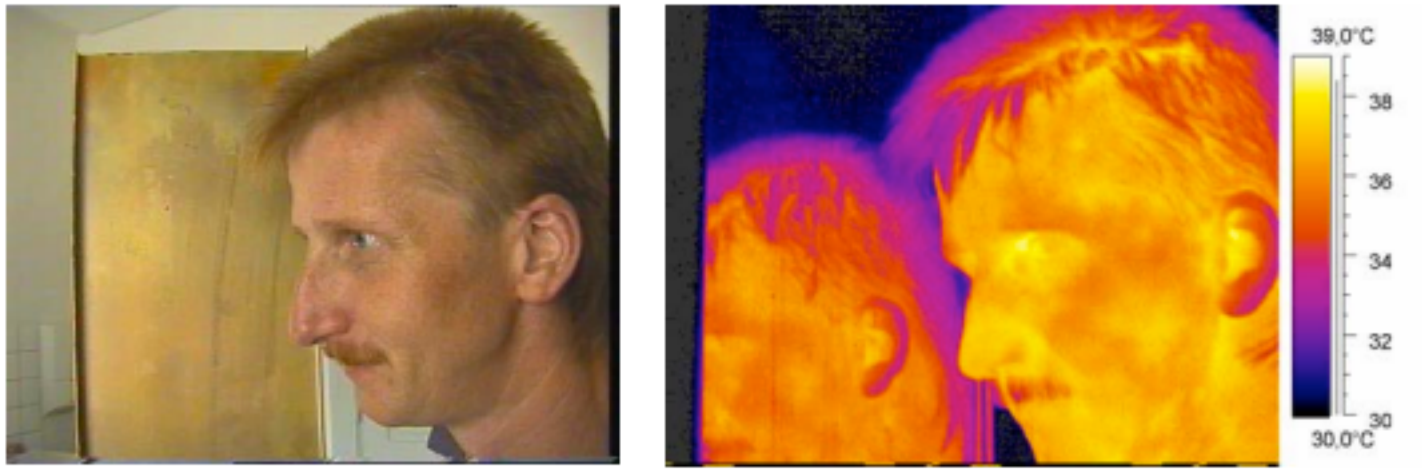


Figure 5. An oxidized old brass plate with a lot of surface roughness in the 1µm scale or below is scattering light diffusely for visible light, but at least in part specularly for thermal IR radiation of $\lambda \approx 10\mu m$.

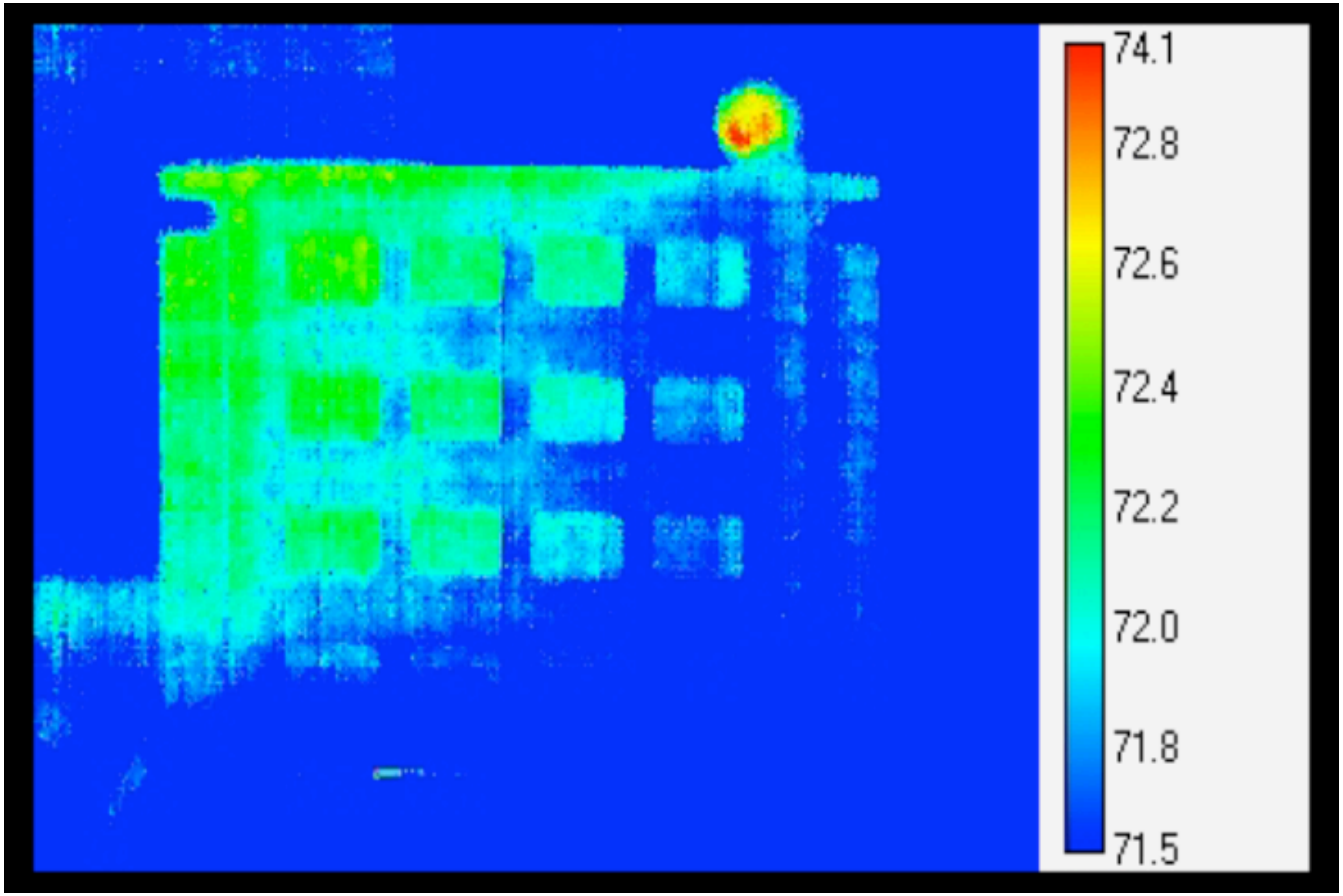
(images from
“Identification and suppression of thermal reflections in infrared thermal imaging,”
Henke et. al.,
InfraMation 2004.)

High conductivity of metal meant residue spread within seconds

So the rest of our results are only for plastic keypads

An ideal run

An ideal run

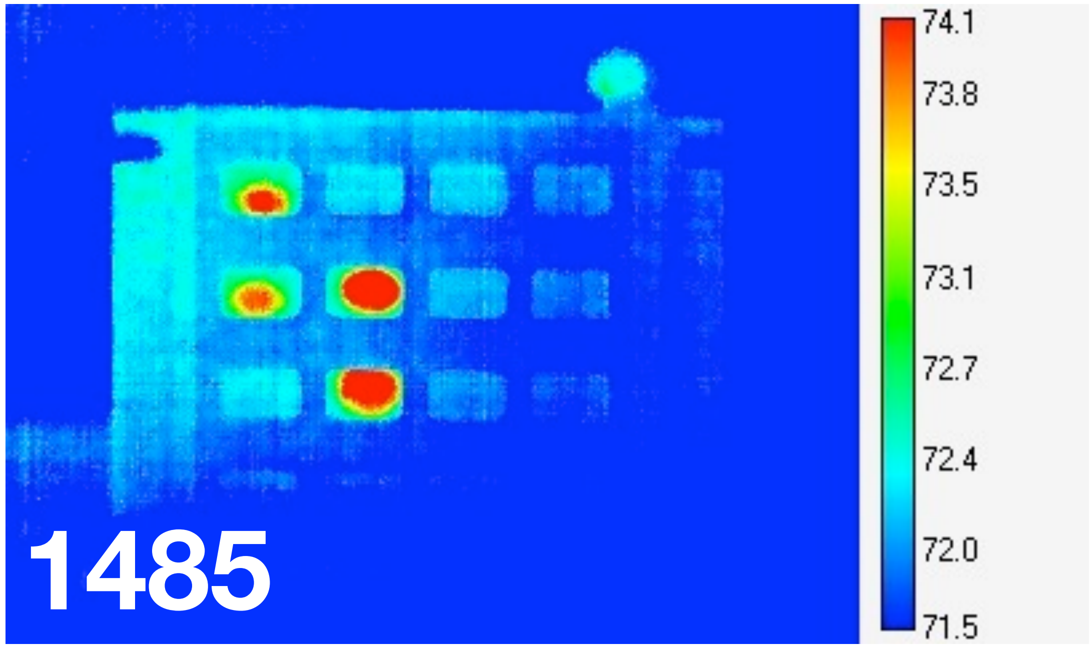


Results can vary widely

Even in the first frame after entry, see very different pictures:

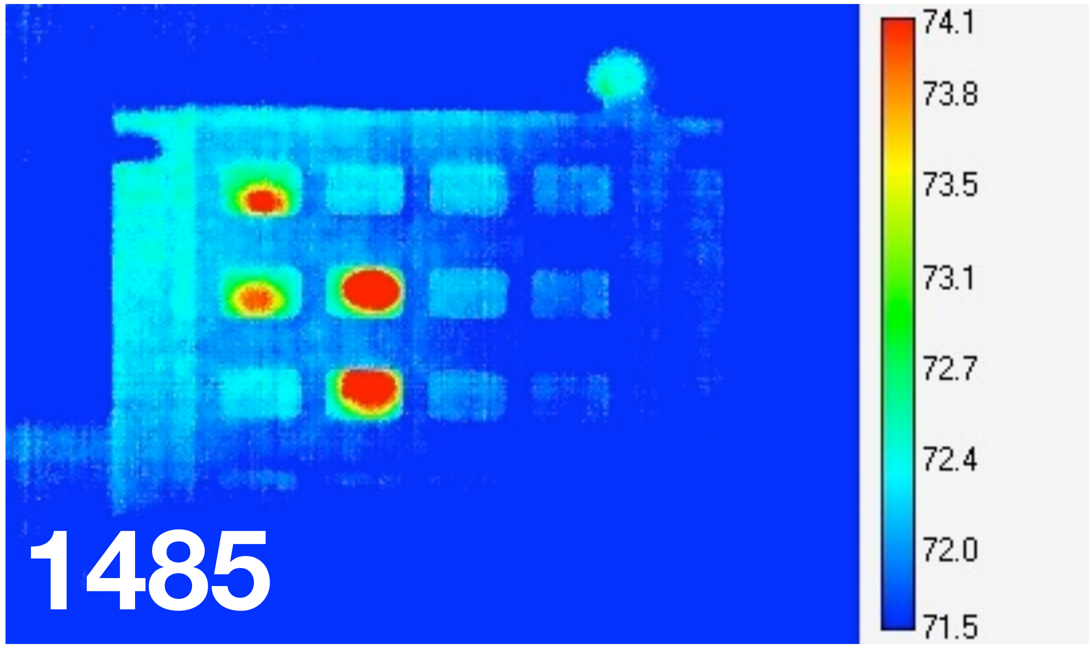
Results can vary widely

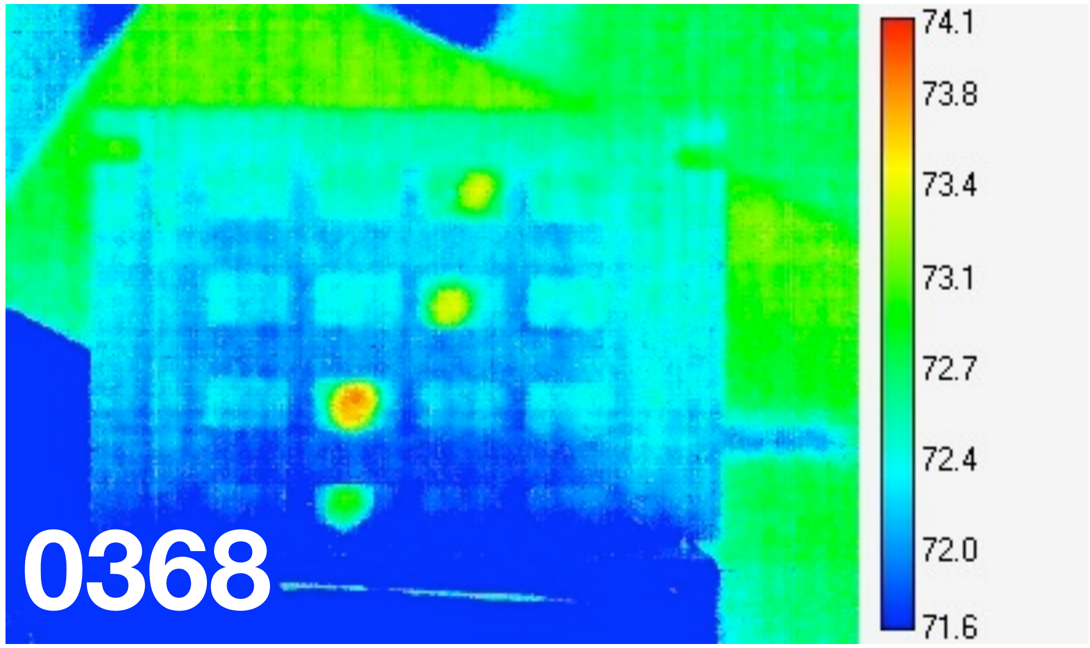
Even in the first frame after entry, see very different pictures:



Results can vary widely

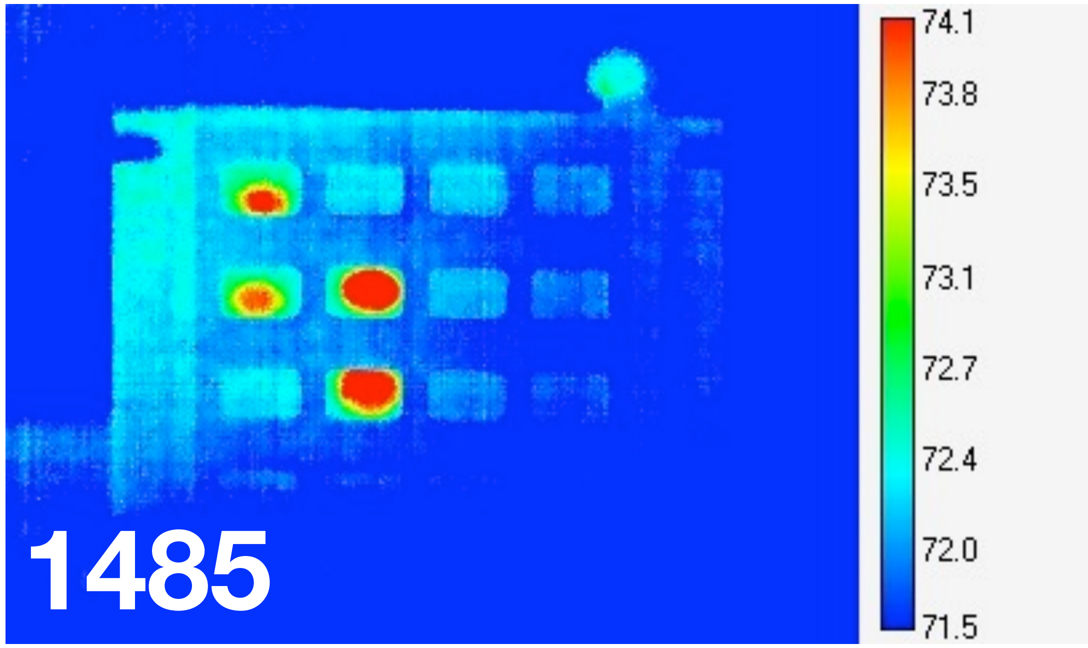
Even in the first frame after entry, see very different pictures:

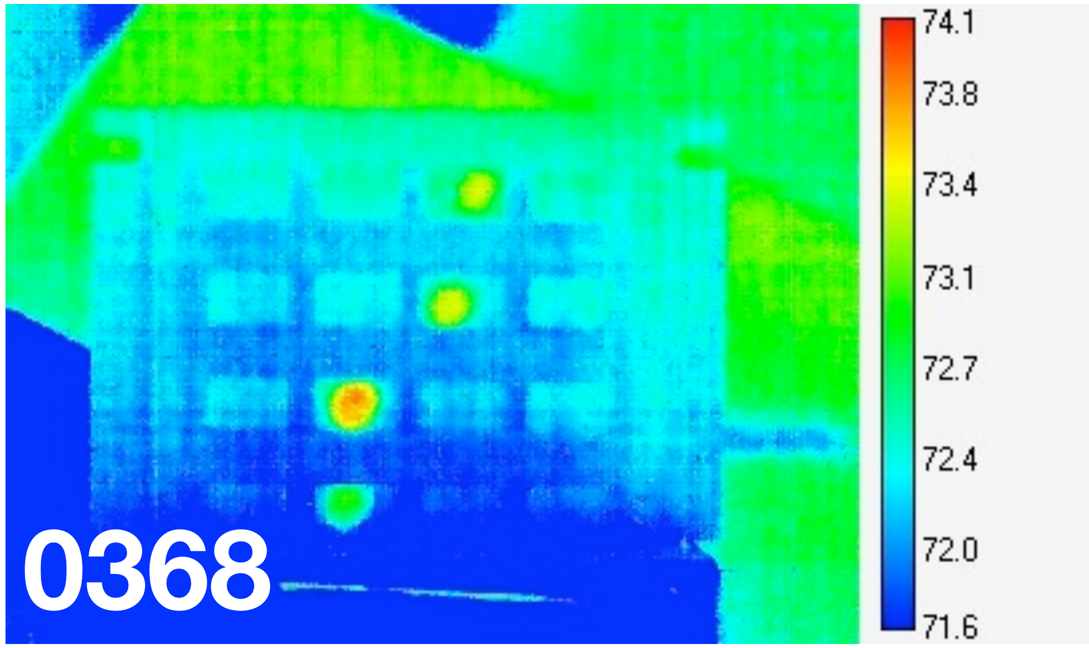


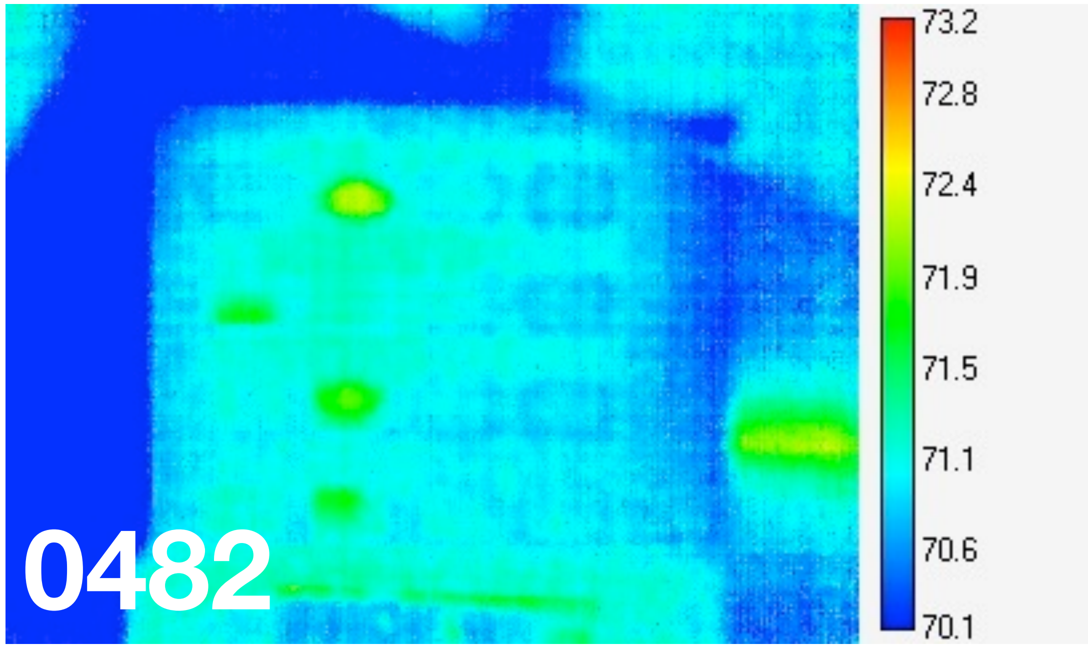


Results can vary widely

Even in the first frame after entry, see very different pictures:

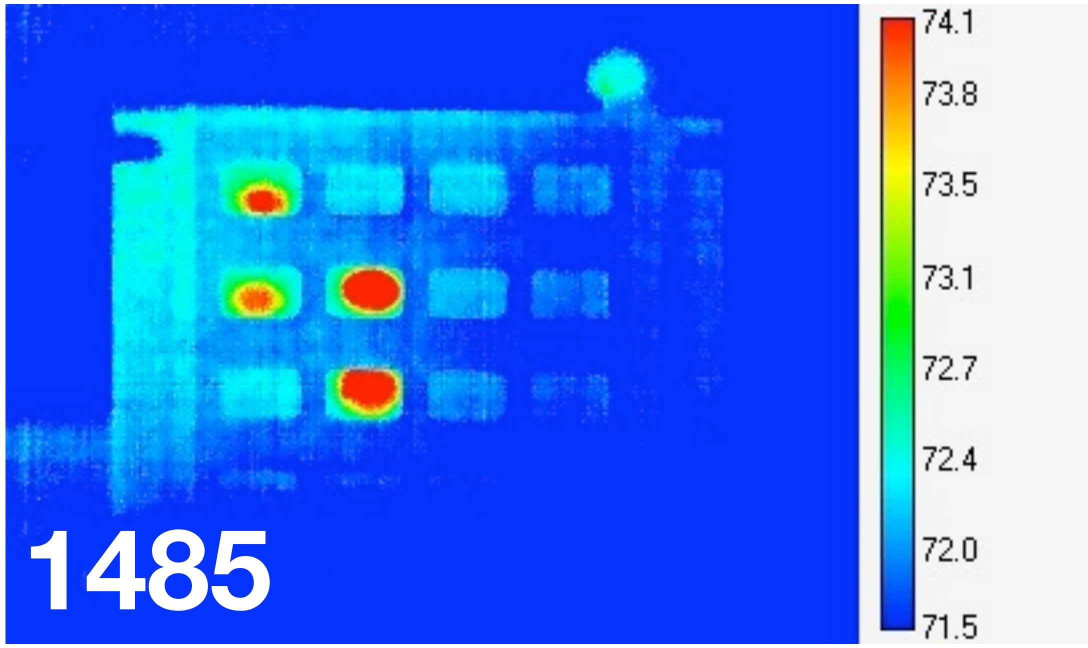


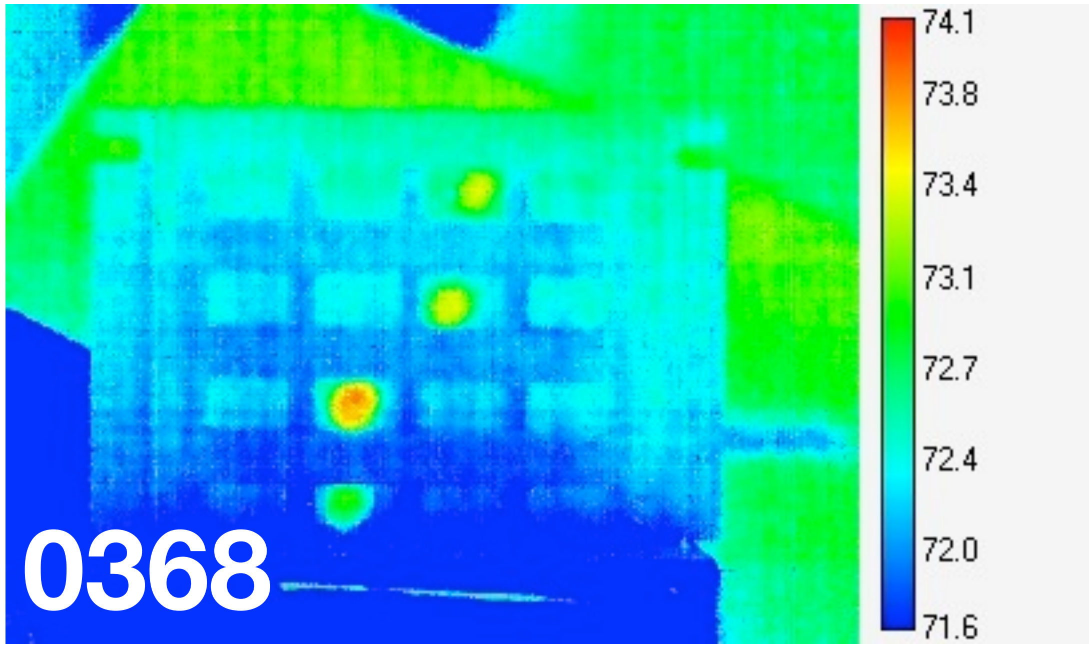


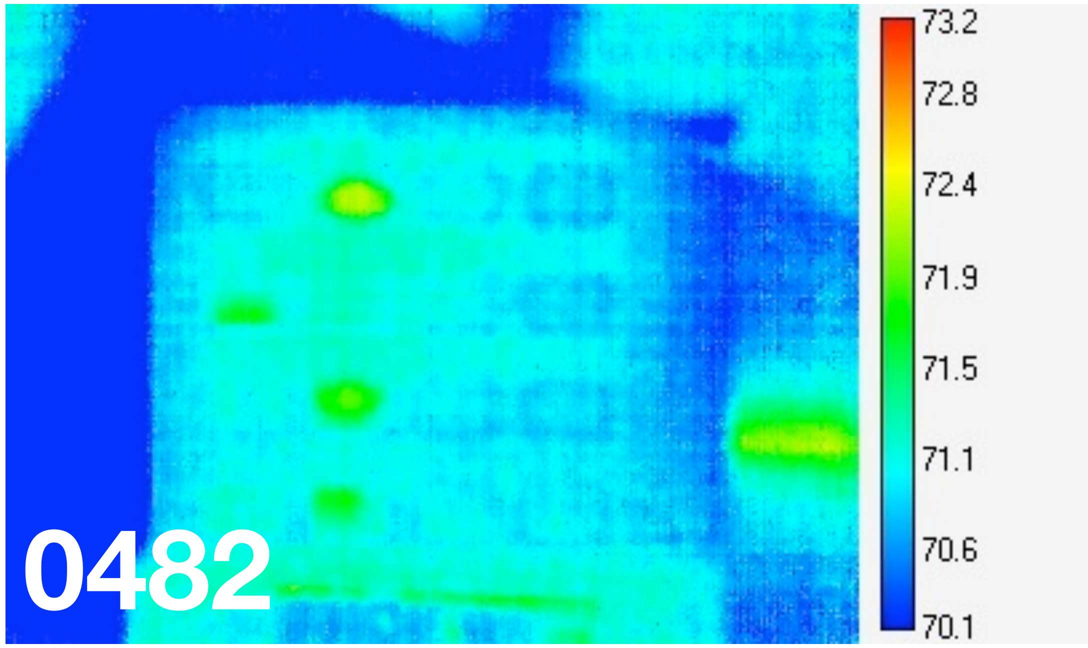


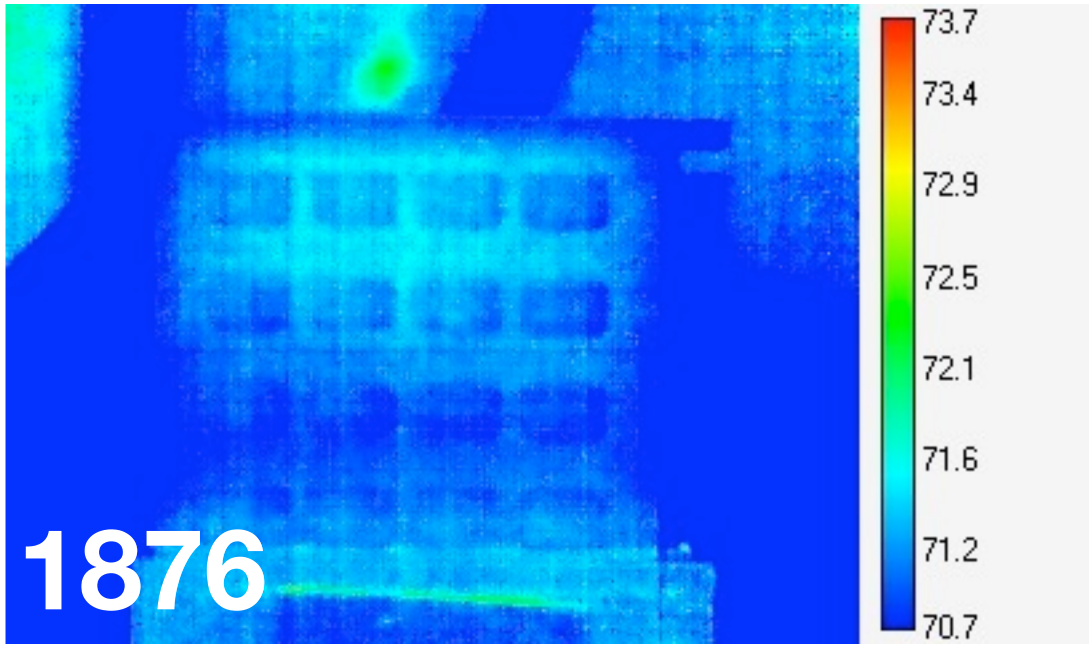
Results can vary widely

Even in the first frame after entry, see very different pictures:







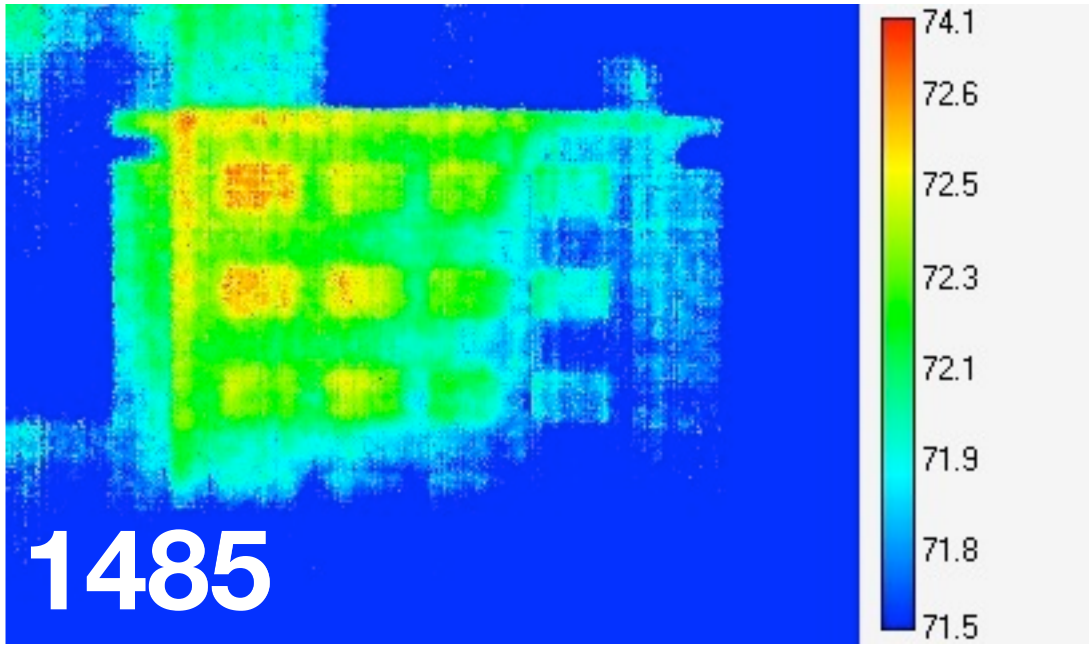


Results can vary widely

See similar differences in how residue degrades over time:

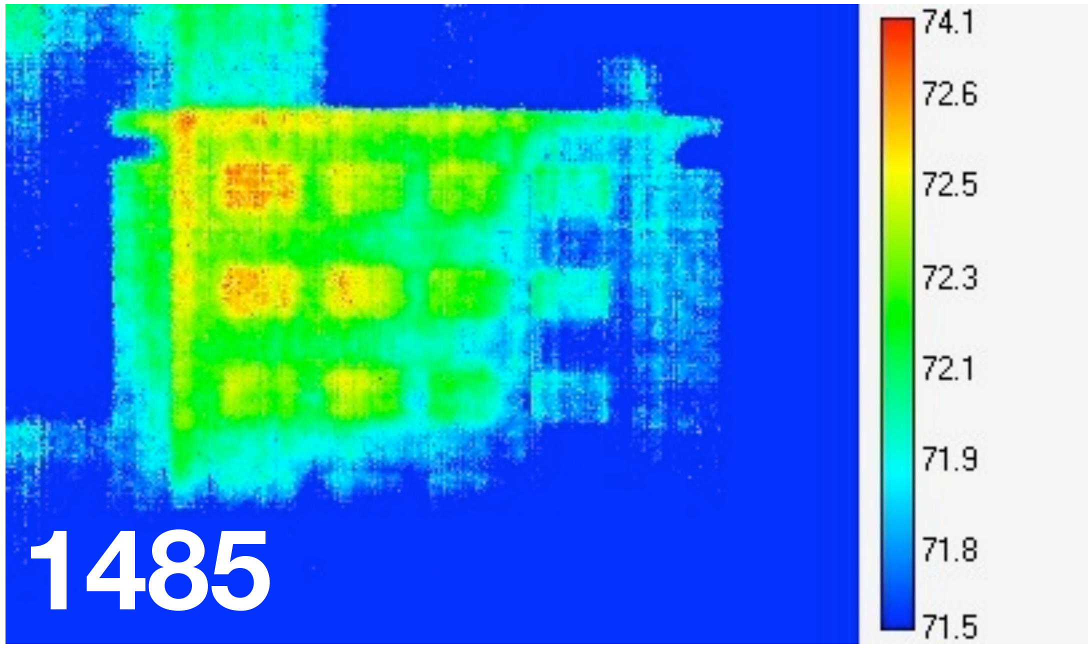
Results can vary widely

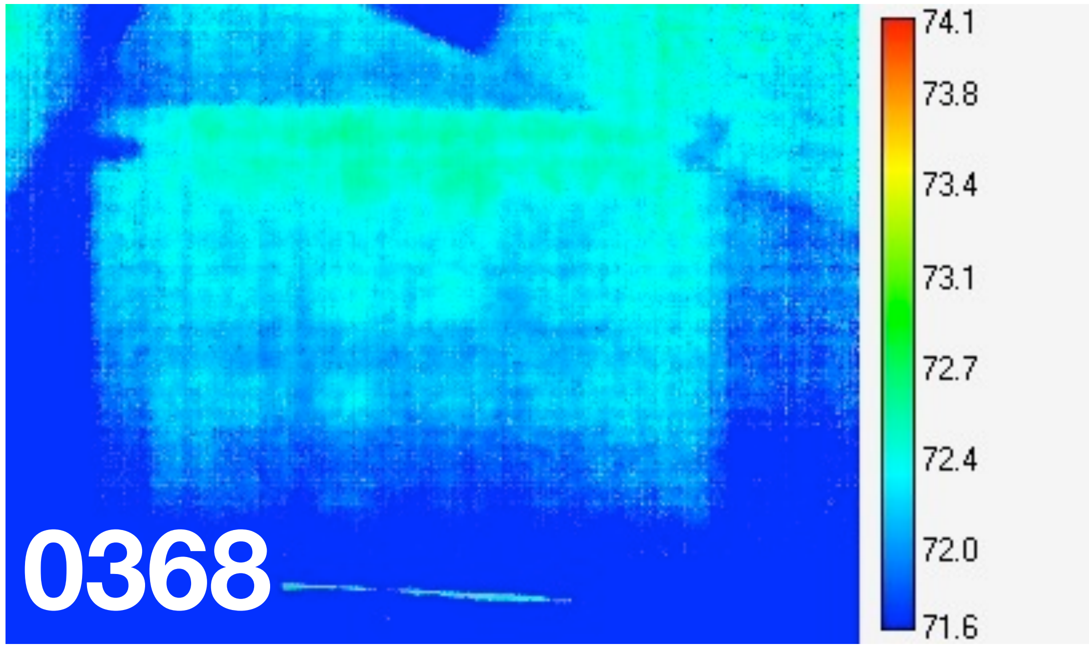
See similar differences in how residue degrades over time:

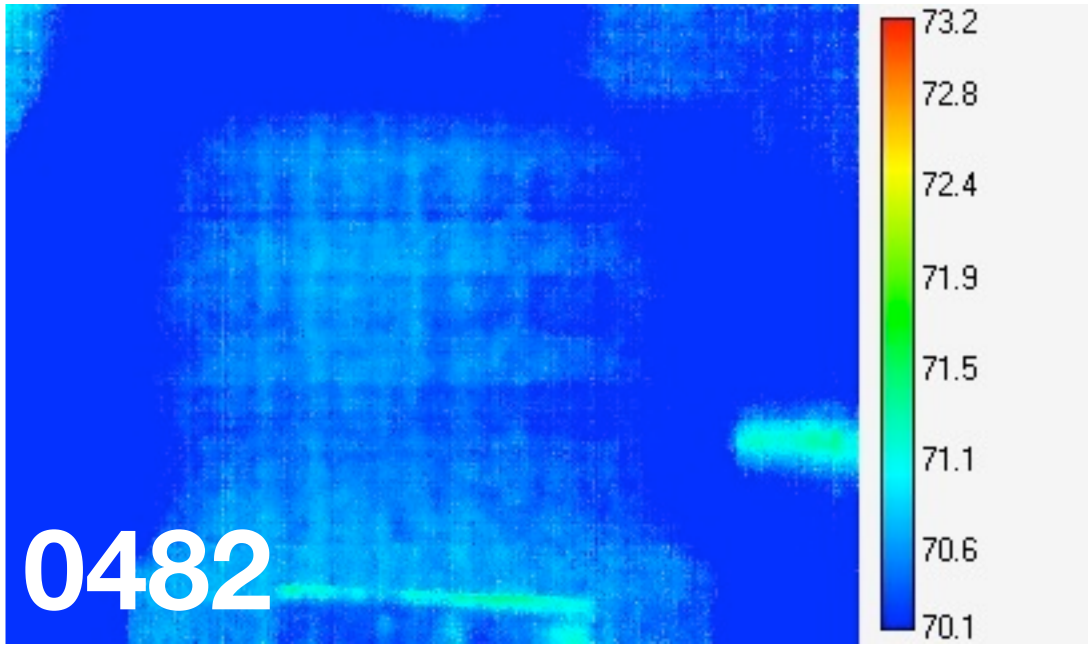


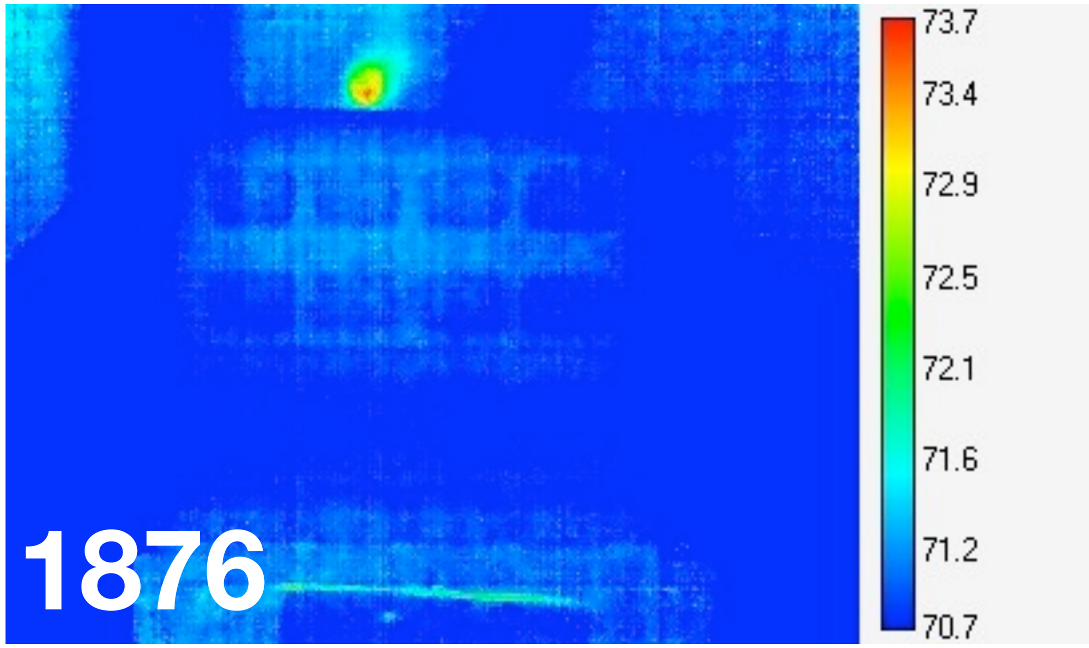
Results can vary widely

See similar differences in how residue degrades over time:









Outline

Experiment design

Camera data

Analyzing the data

Conclusions

Human review

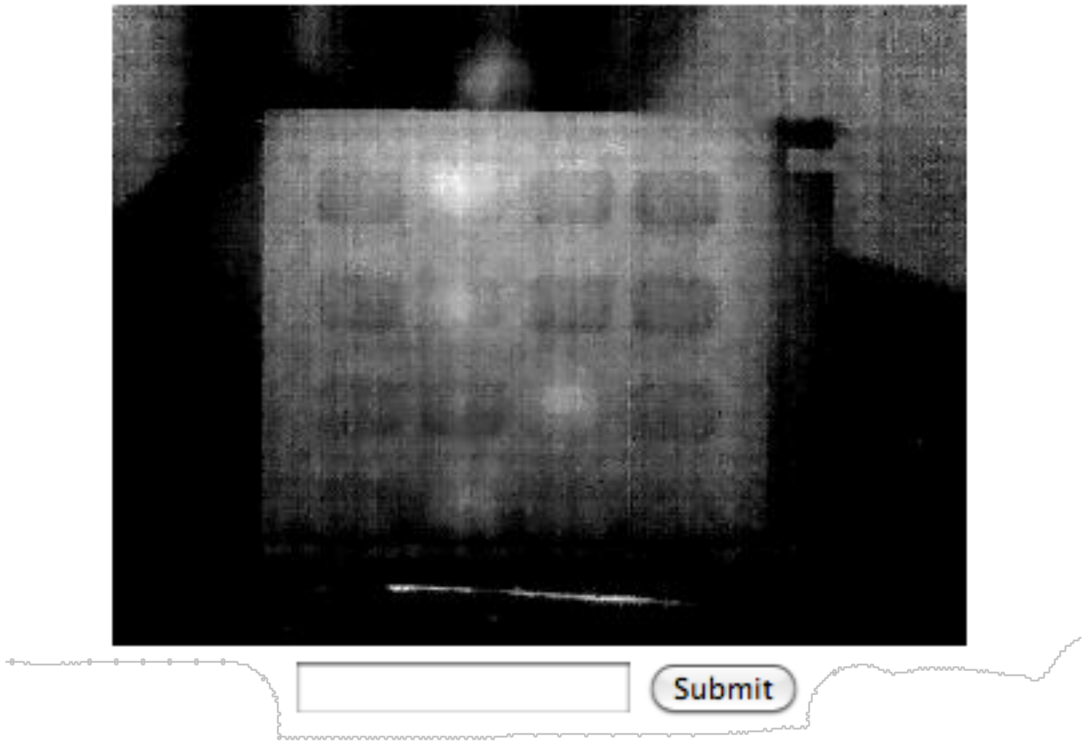
Human review

First approach: human visual inspection

Human review

First approach: human visual inspection

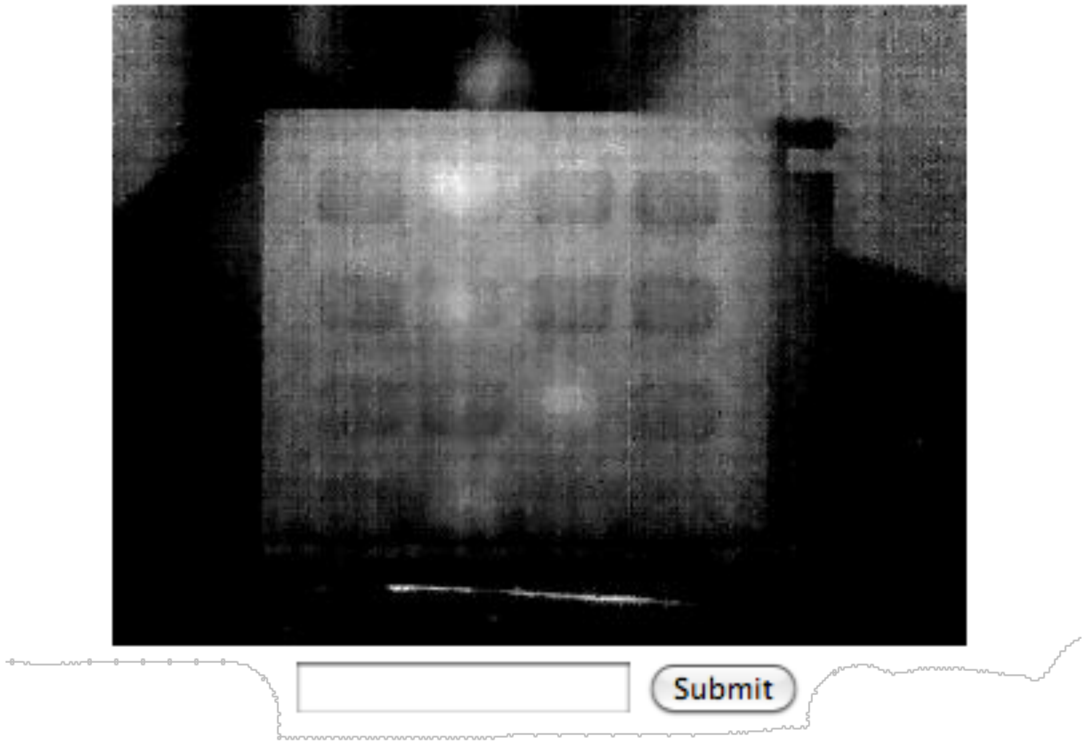
- Examine every 10th frame (in random order) to guess code entered



Human review

First approach: human visual inspection

- Examine every 10th frame (in random order) to guess code entered

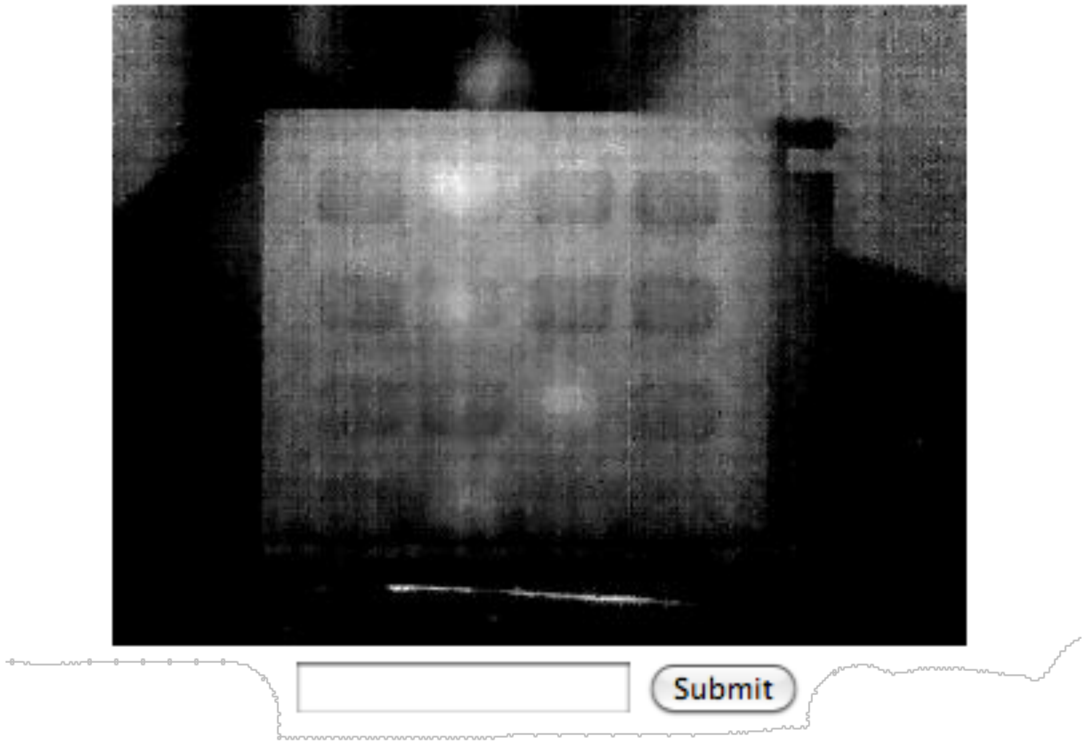


Problem: this approach doesn't scale very well! (looked at ~1800 images)

Human review

First approach: human visual inspection

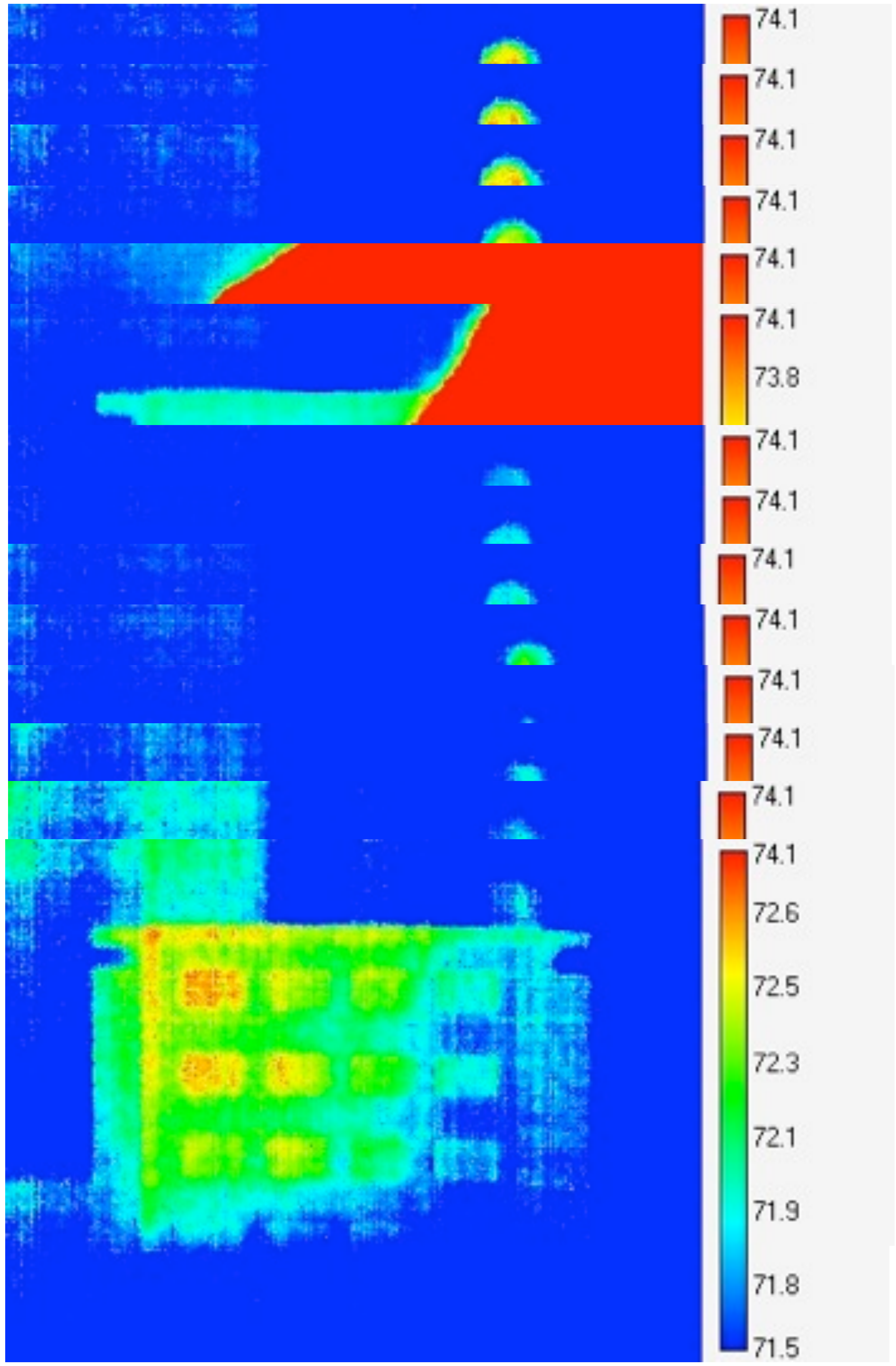
- Examine every 10th frame (in random order) to guess code entered



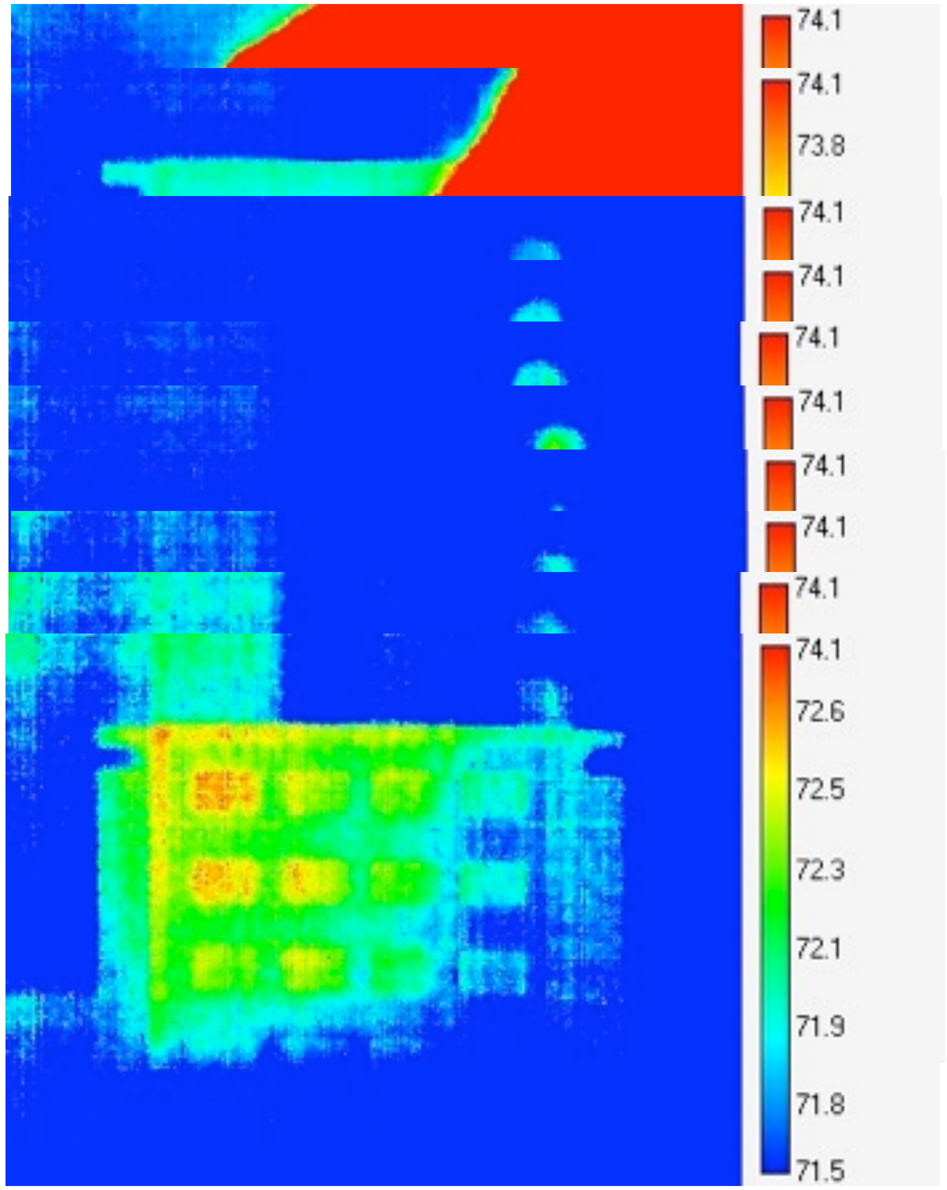
Problem: this approach doesn't scale very well! (looked at ~1800 images)

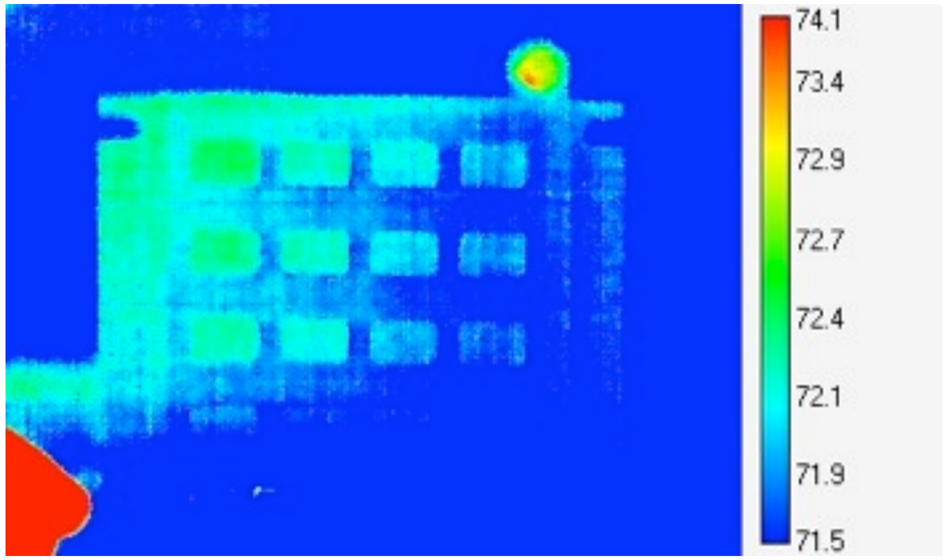
- Second approach: automated review

Automated review: what to do with all this footage?

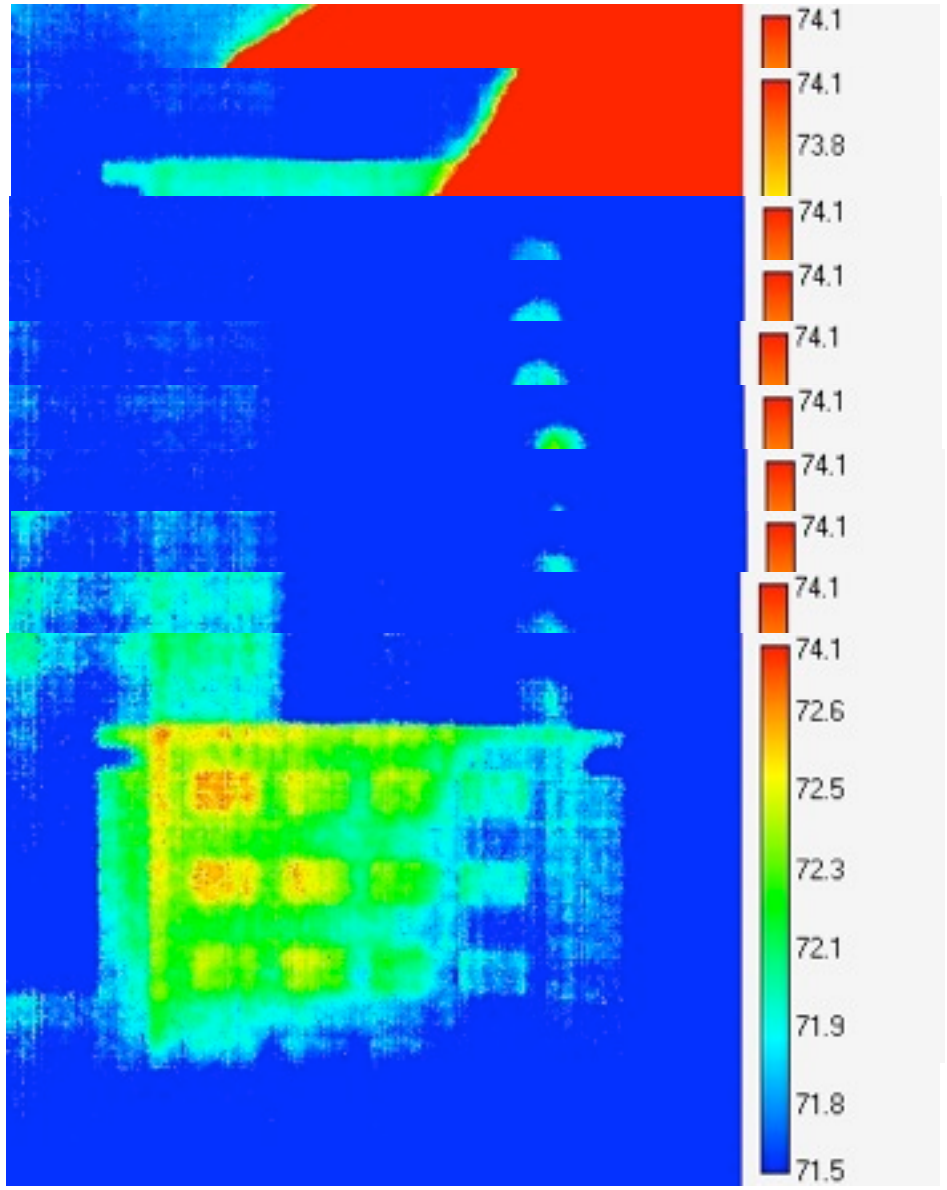


Automated review: what to do with all this footage?

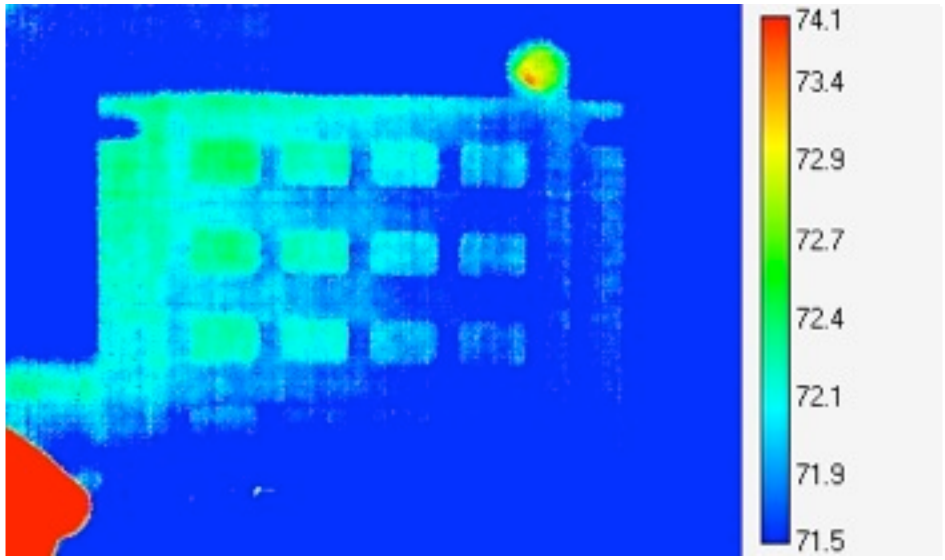




Automated review: what to do with all this footage?

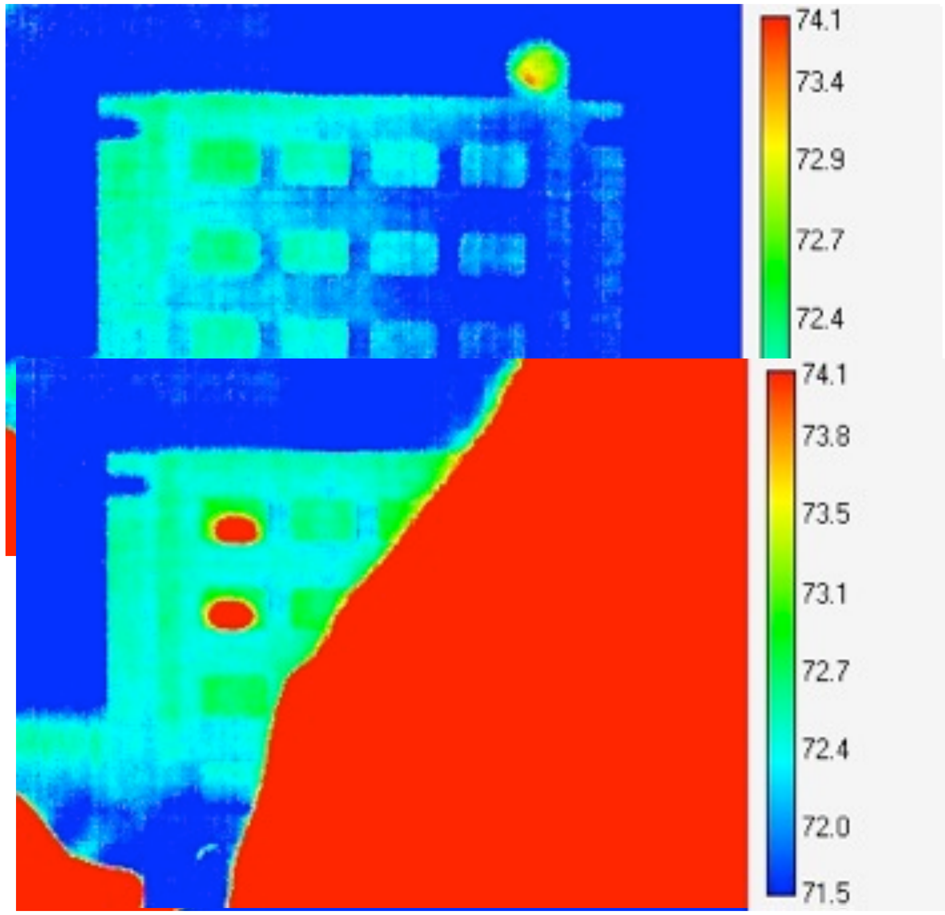


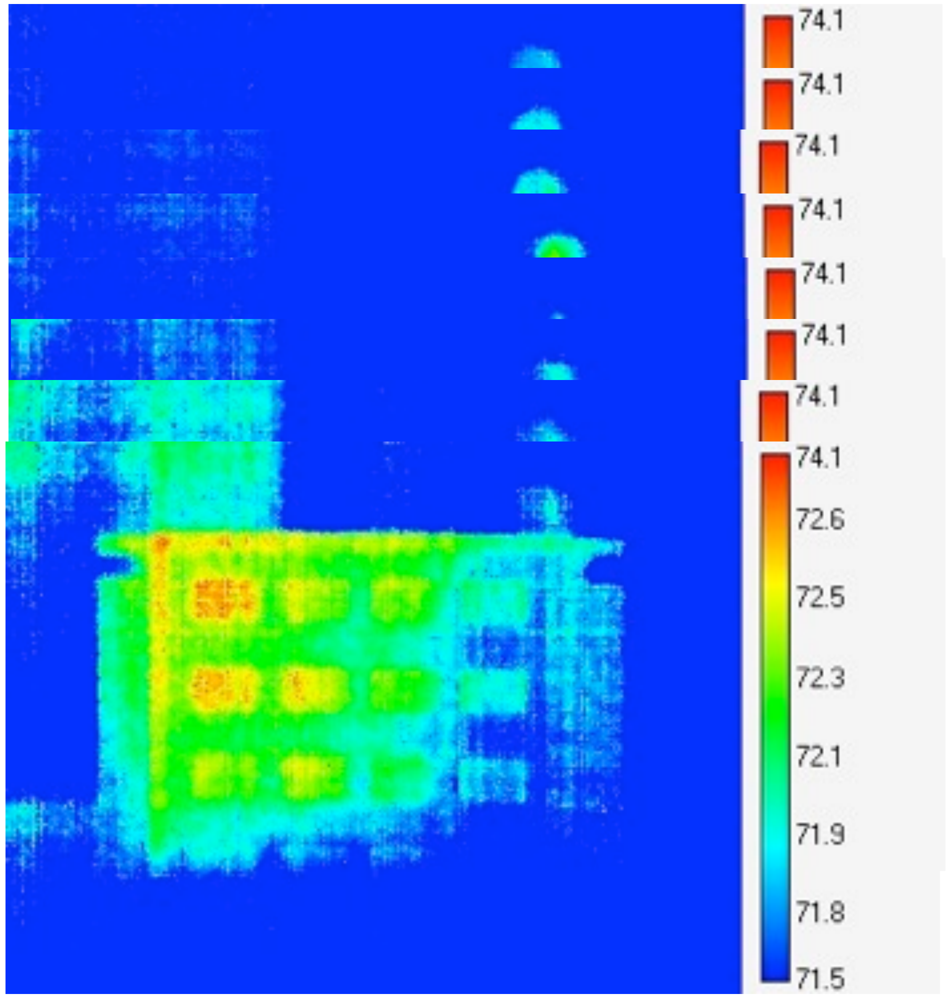
calibration



Automated review: what to do with all this footage?

calibration

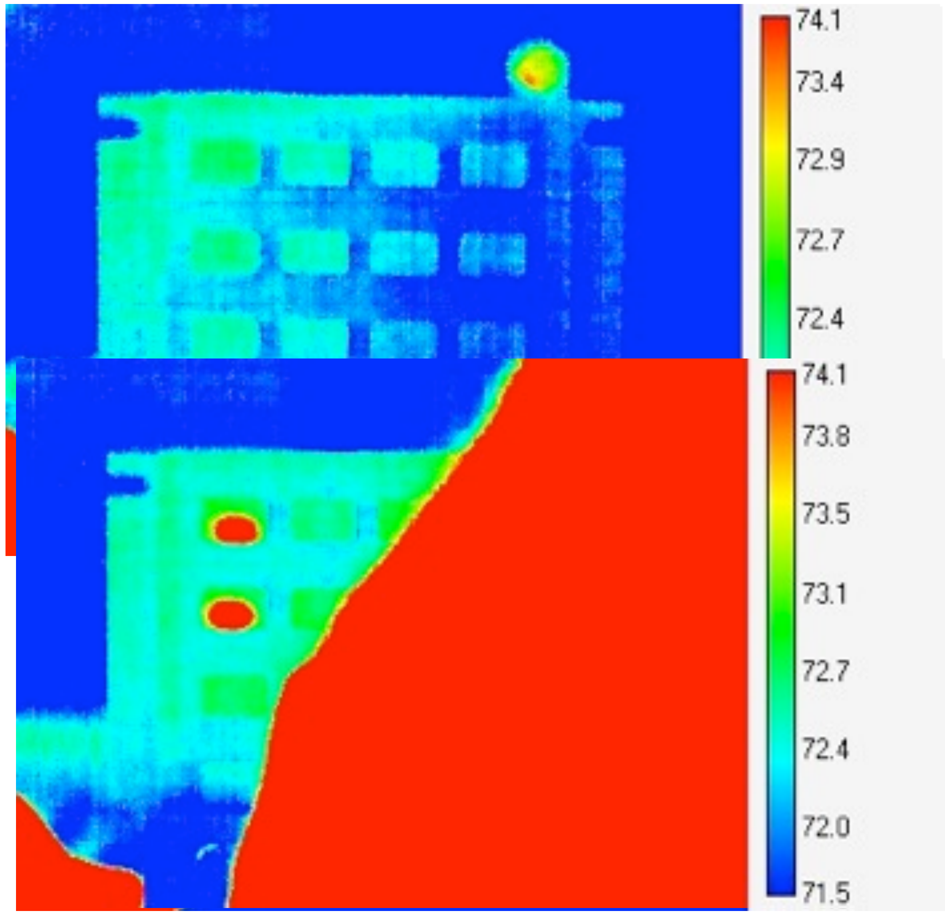


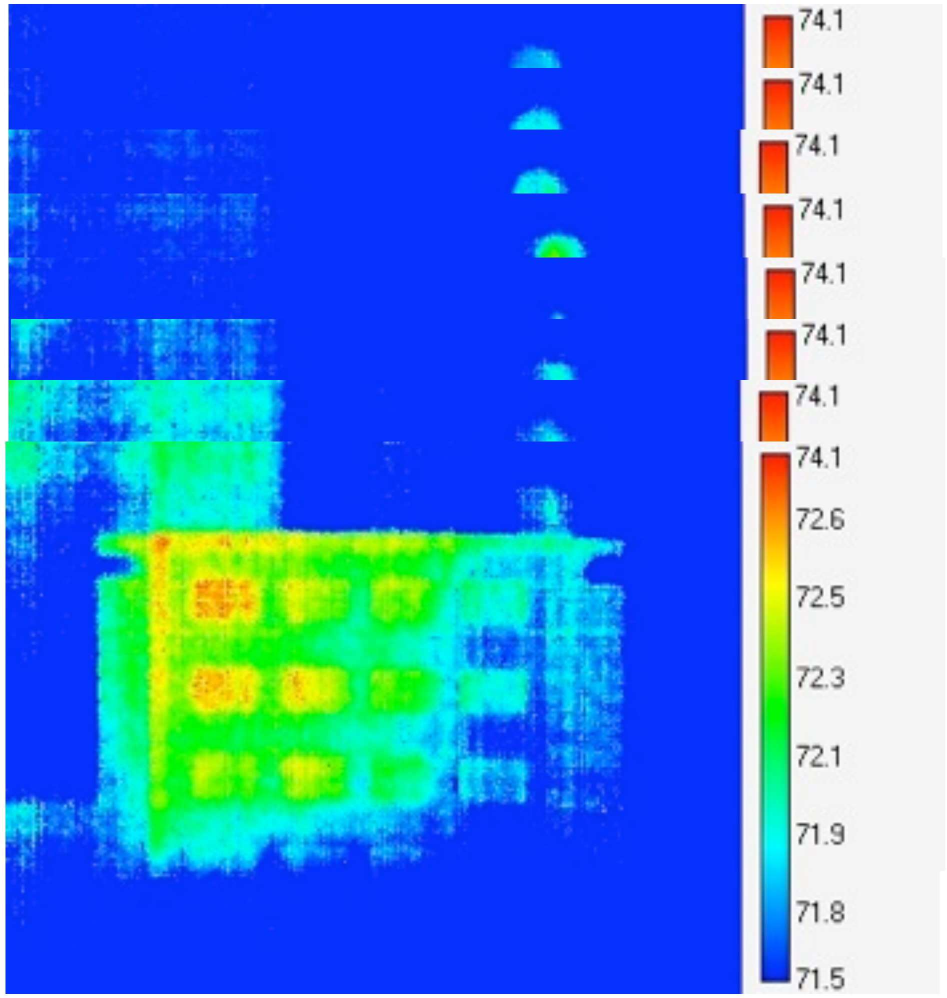


Automated review: what to do with all this footage?

calibration

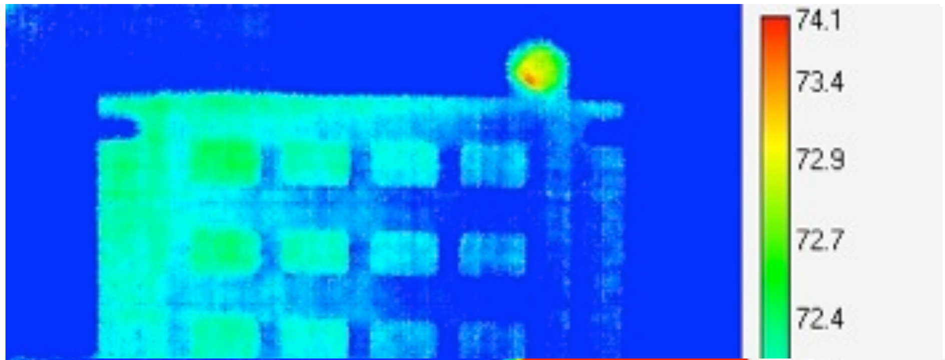
hand



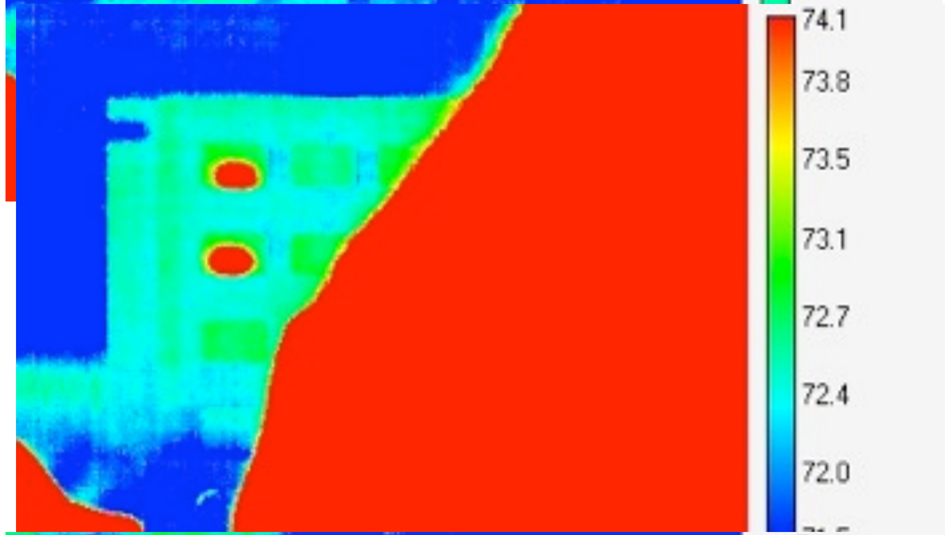


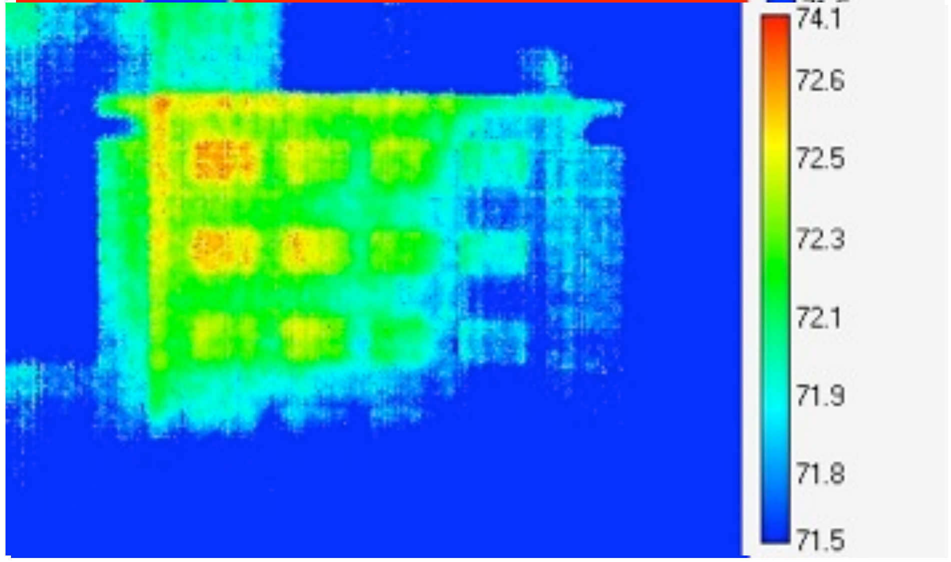
Automated review: what to do with all this footage?

calibration



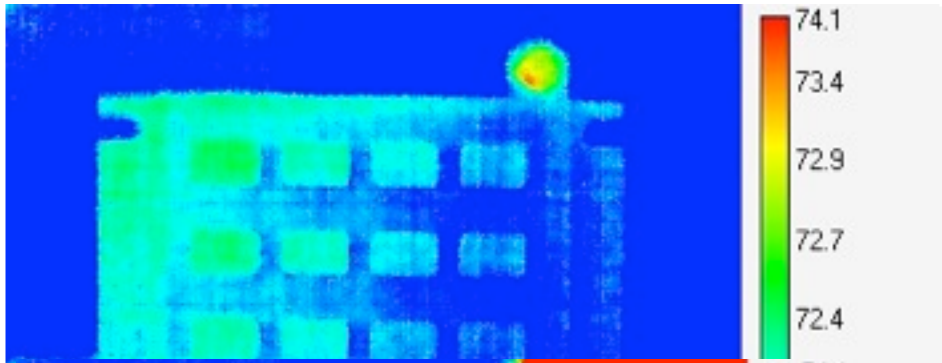
hand



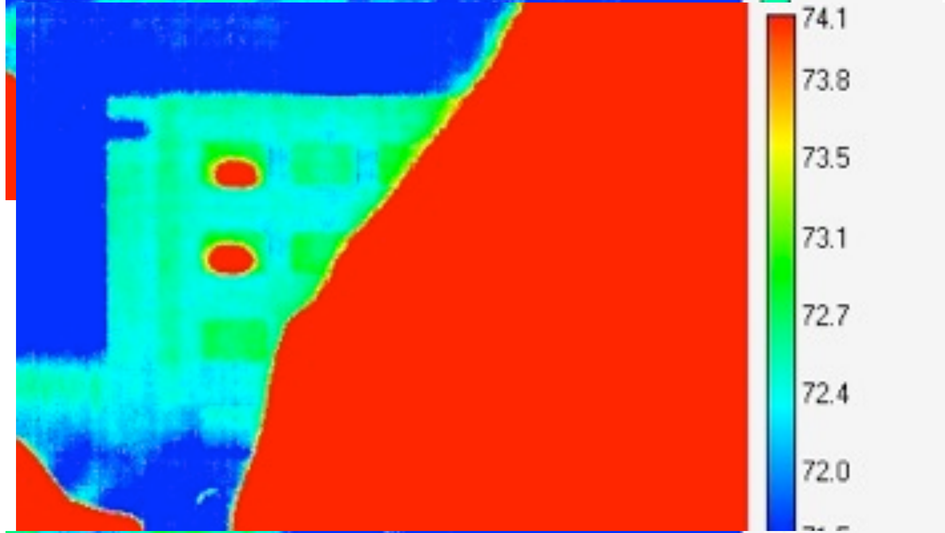


Automated review: what to do with all this footage?

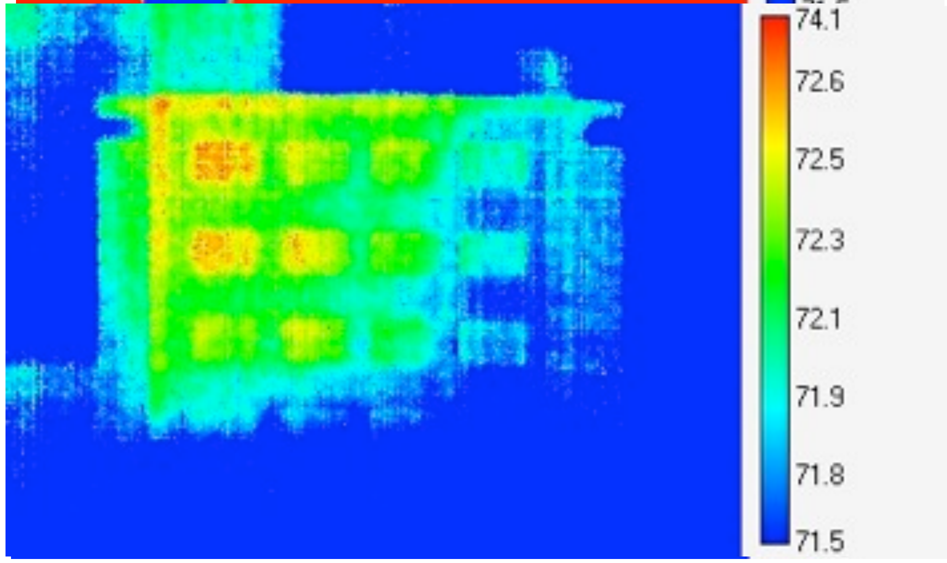
calibration



hand

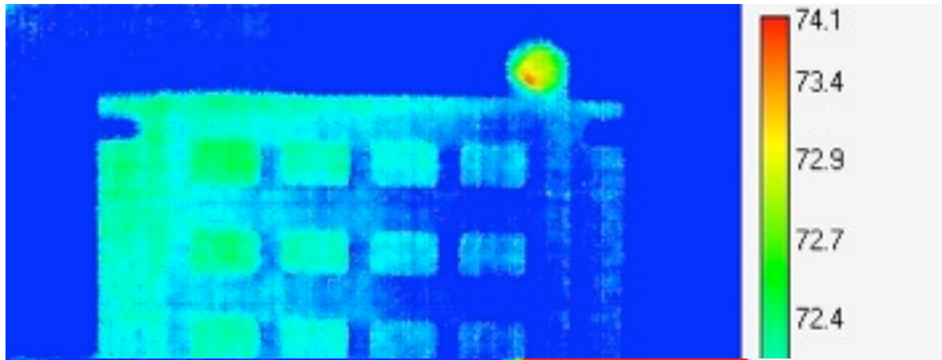


after entry

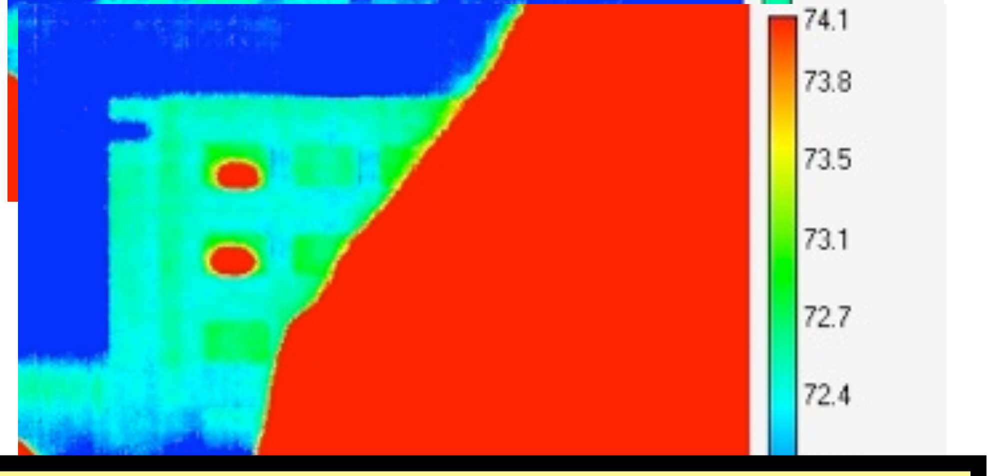


Automated review: what to do with all this footage?

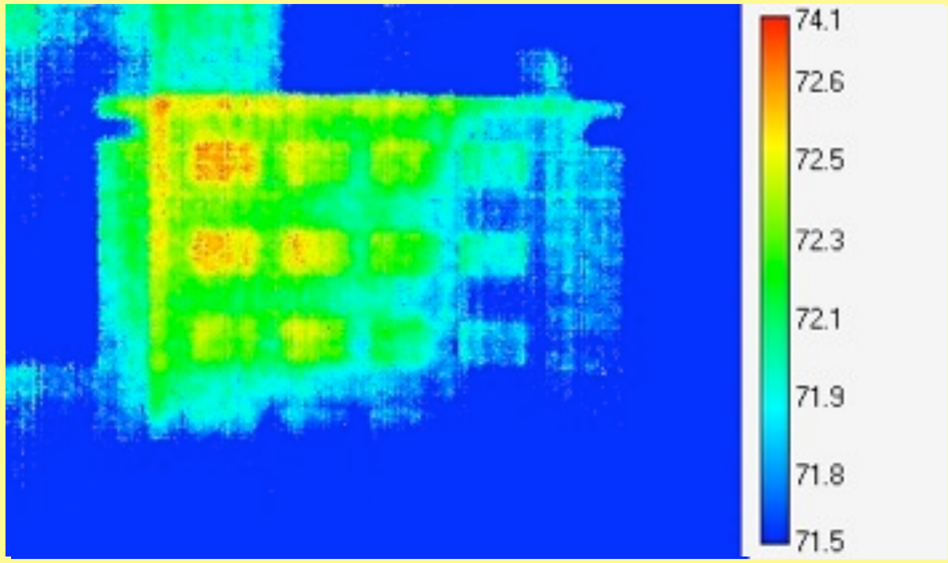
calibration



hand

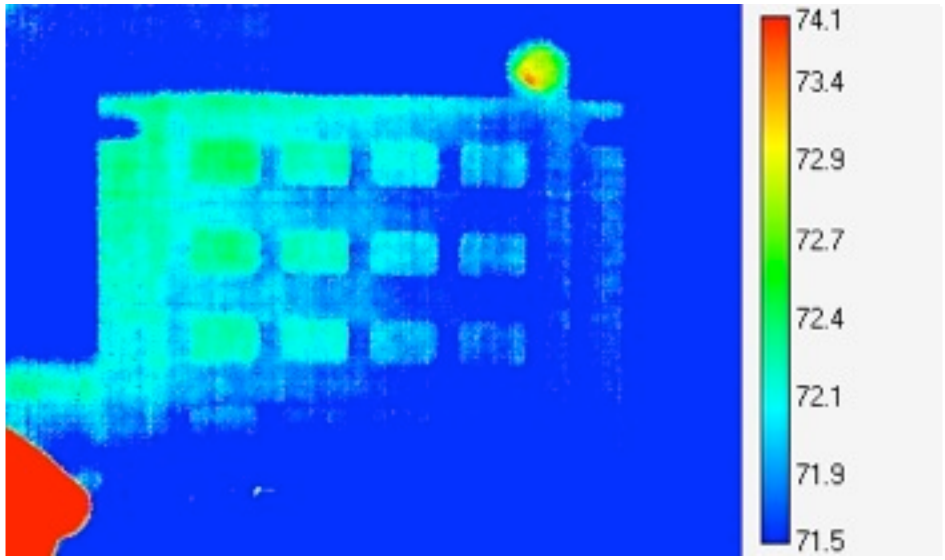


after entry

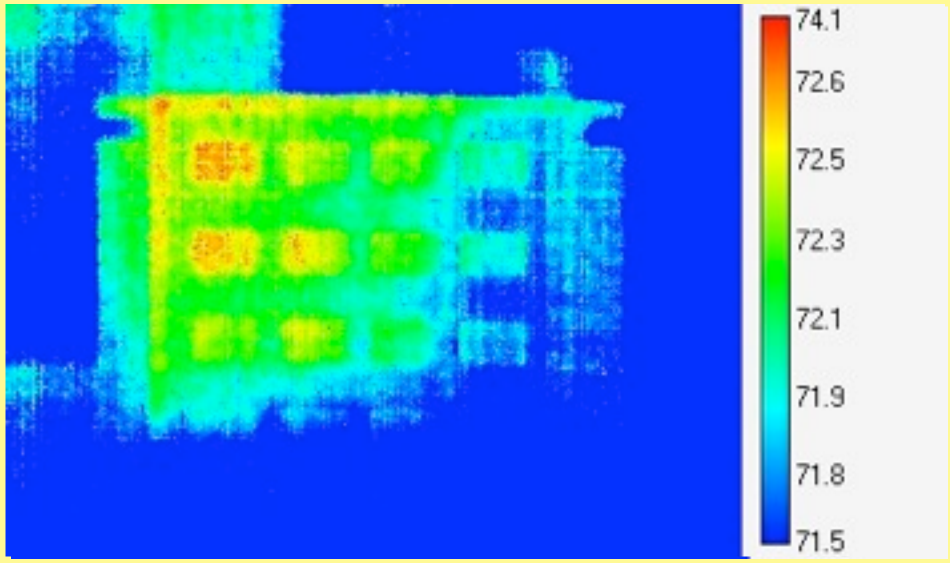


Automated review: what to do with all this footage?

calibration



after entry



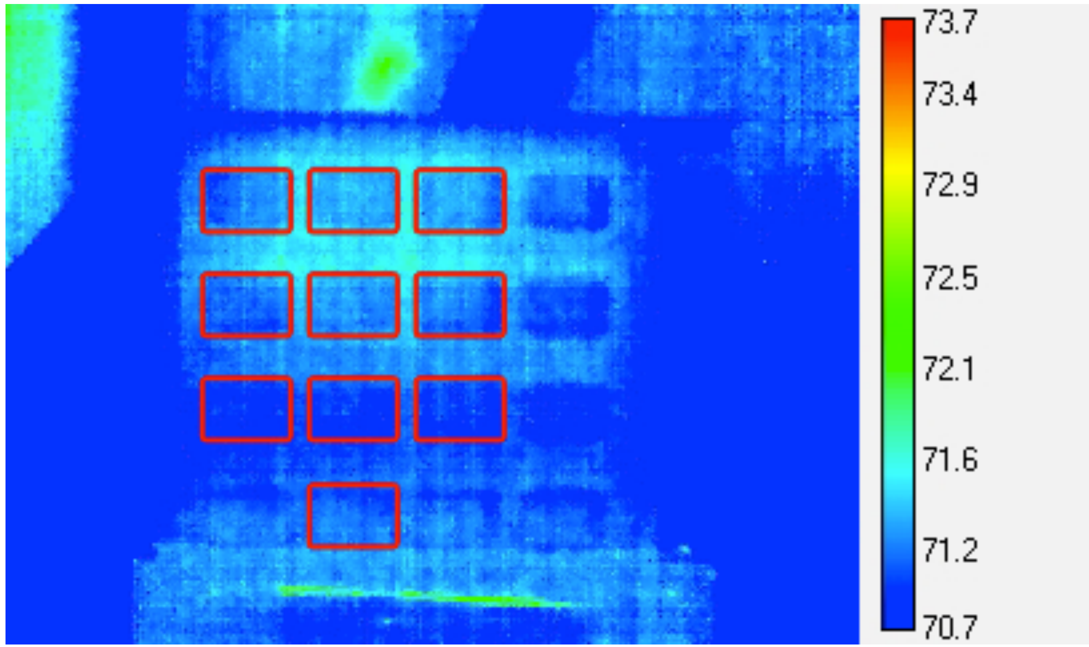
Automated review: which buttons were pressed?

Automated review: which buttons were pressed?

Basic idea: for each region, determine if it is hot above a certain threshold

Automated review: which buttons were pressed?

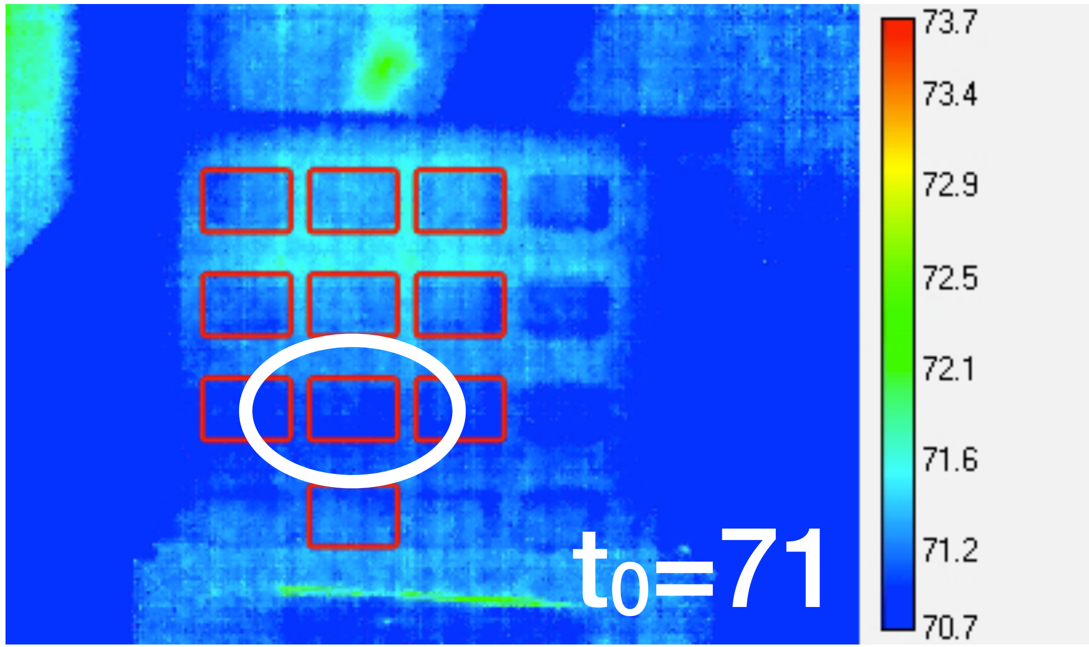
Basic idea: for each region, determine if it is hot above a certain threshold



calibration

Automated review: which buttons were pressed?

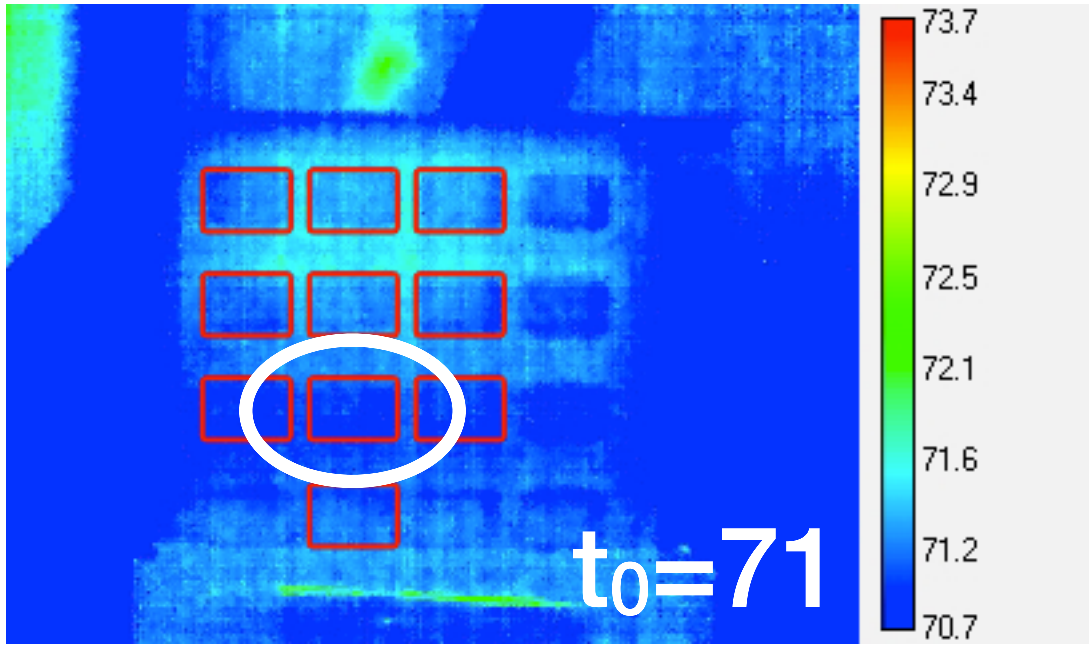
Basic idea: for each region, determine if it is hot above a certain threshold



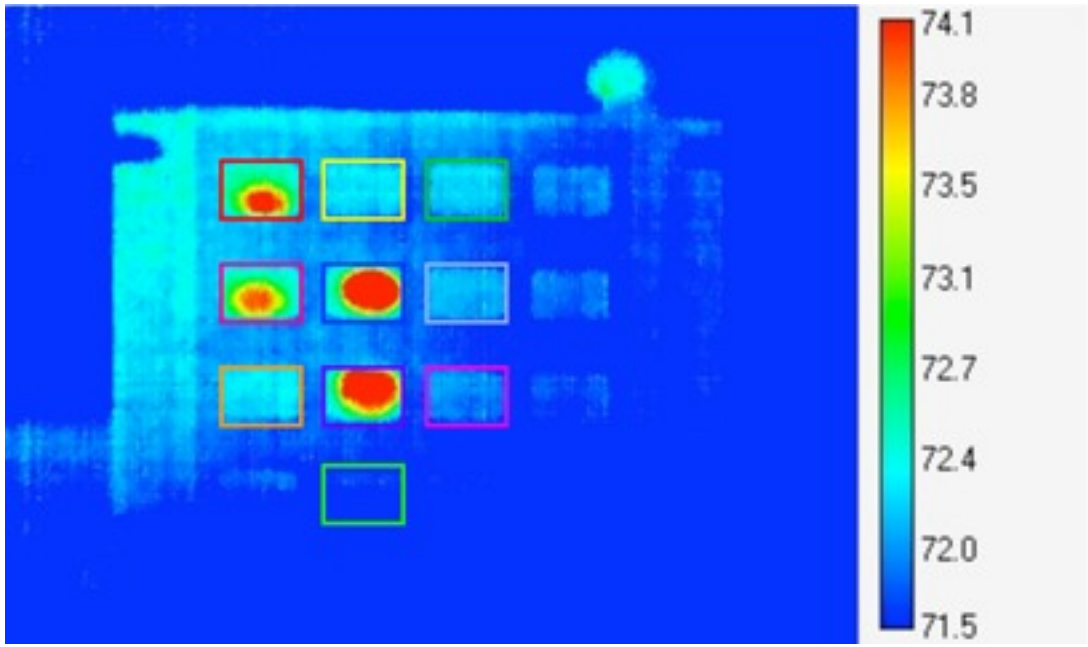
calibration

Automated review: which buttons were pressed?

Basic idea: for each region, determine if it is hot above a certain threshold



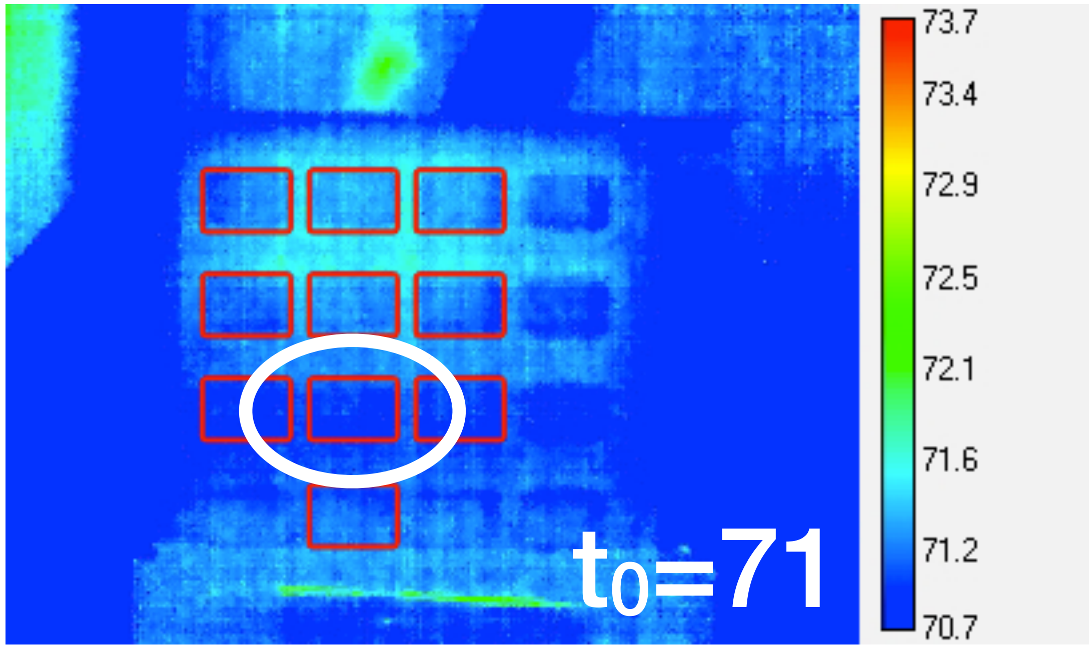
calibration



after entry

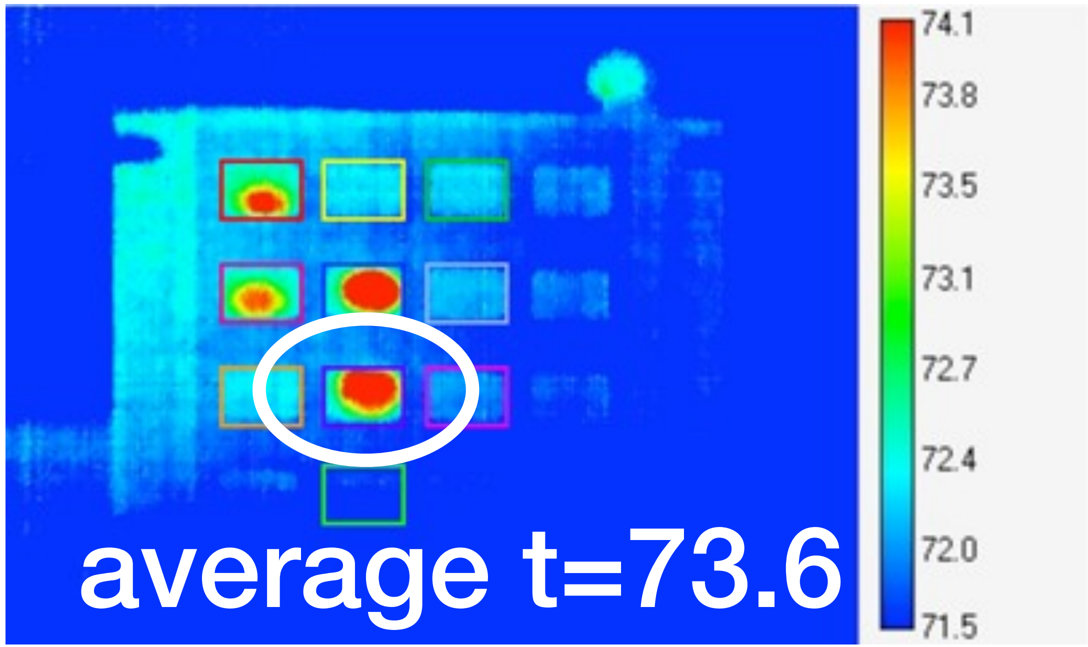
Automated review: which buttons were pressed?

Basic idea: for each region, determine if it is hot above a certain threshold



t_0=71

calibration

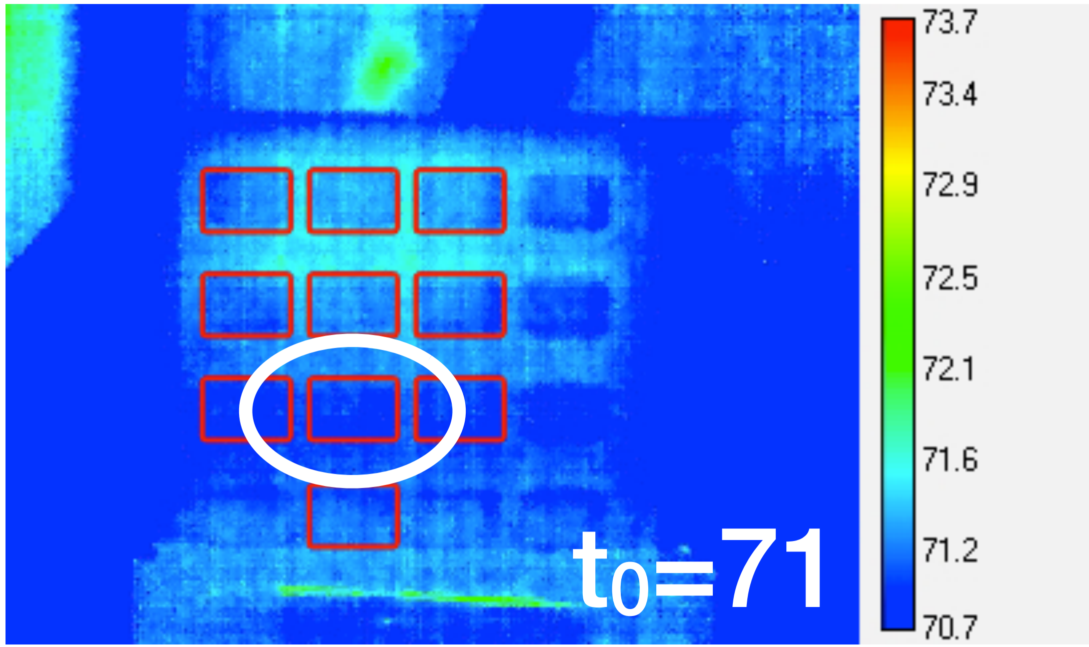


average t=73.6

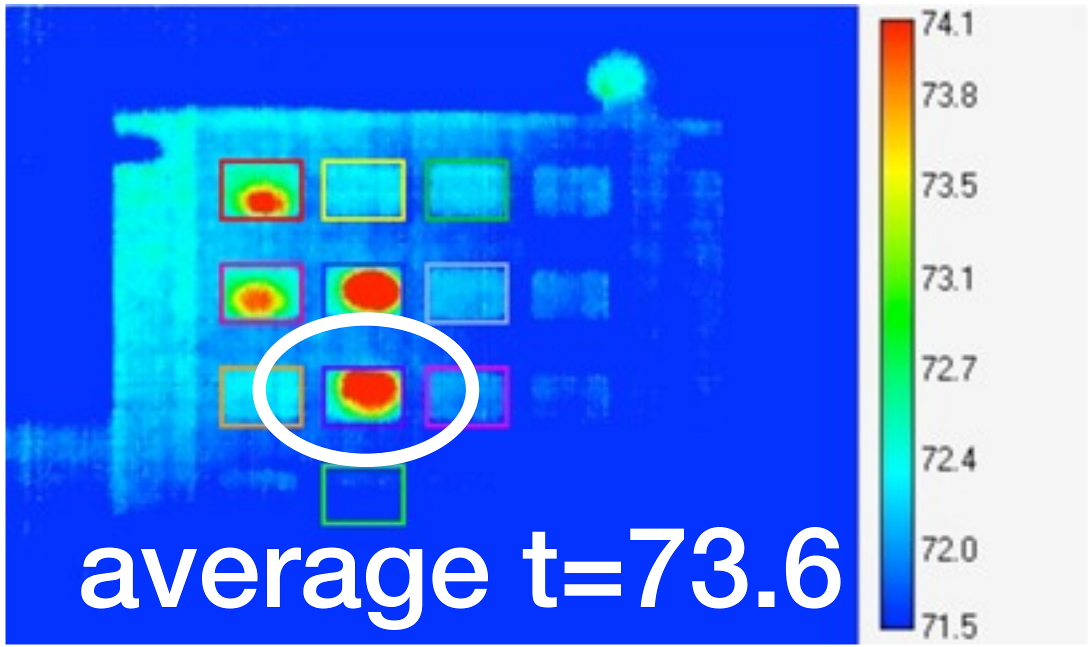
after entry

Automated review: which buttons were pressed?

Basic idea: for each region, determine if it is hot above a certain threshold



calibration

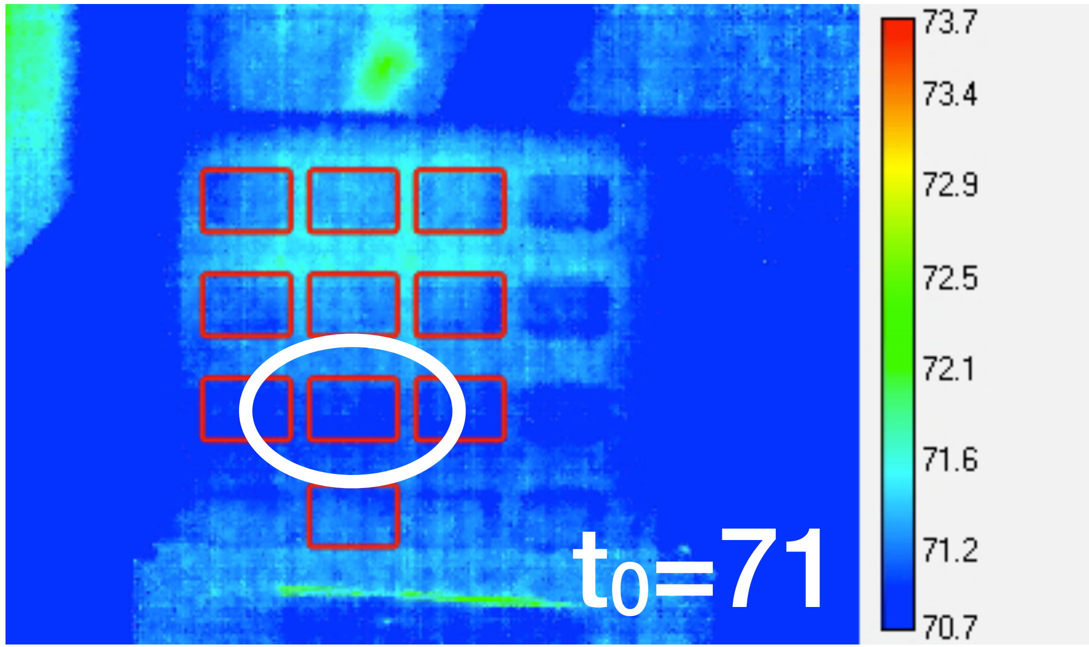


after entry

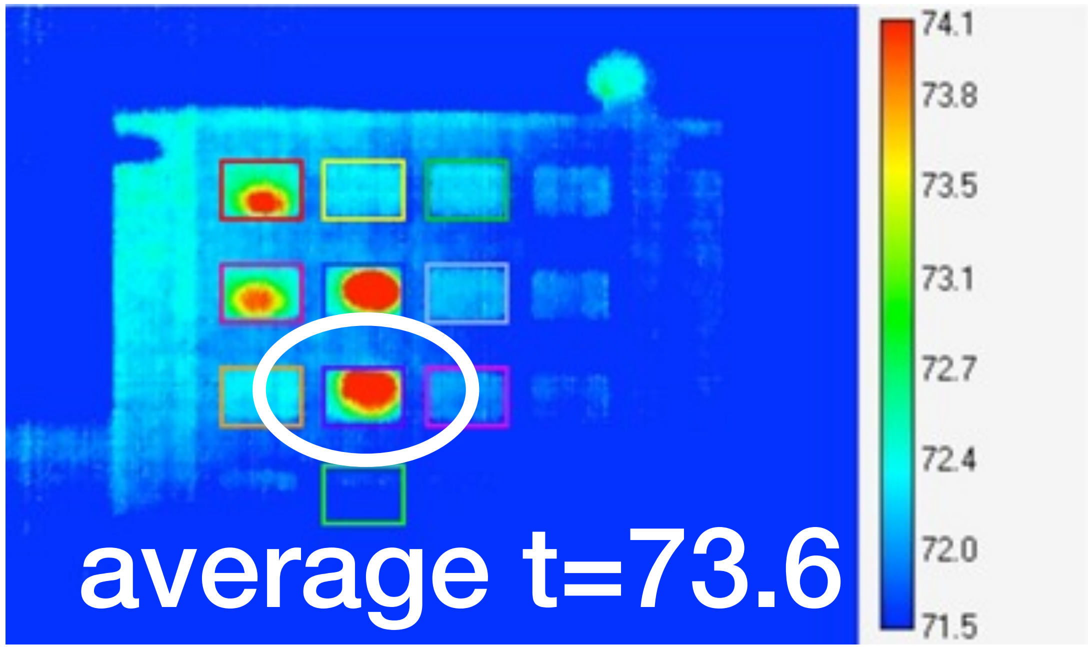
Can repeat this process for each region, then sort in order of $\Delta = t - t_0$

Automated review: which buttons were pressed?

Basic idea: for each region, determine if it is hot above a certain threshold



calibration



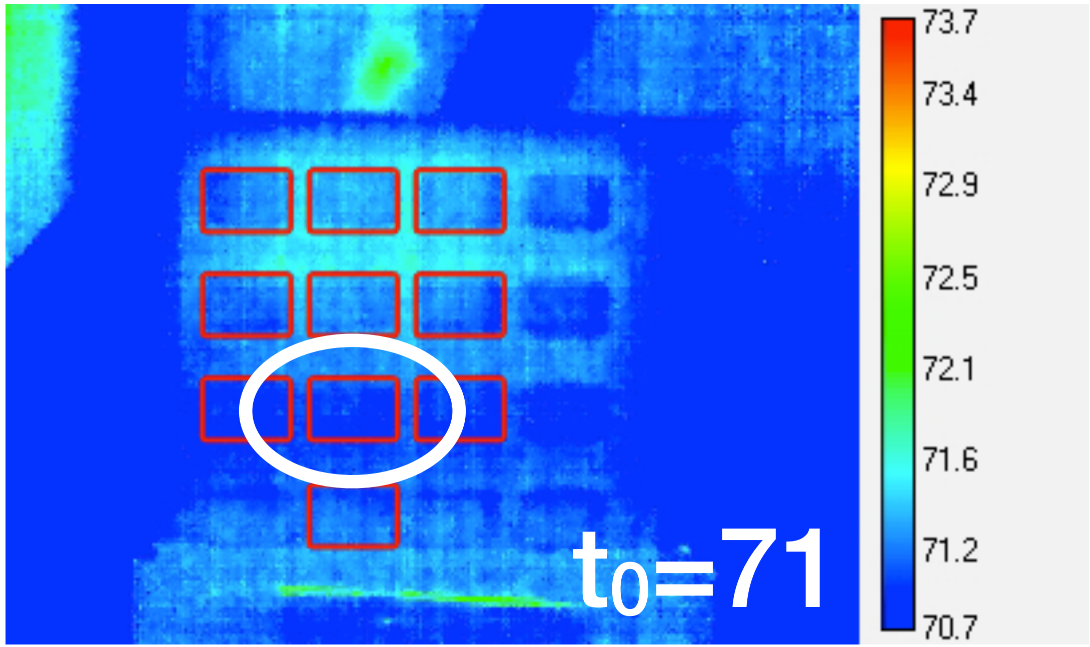
after entry

Can repeat this process for each region, then sort in order of $\Delta = t - t_0$

Examined regions in isolation because we didn't observe much heat spread

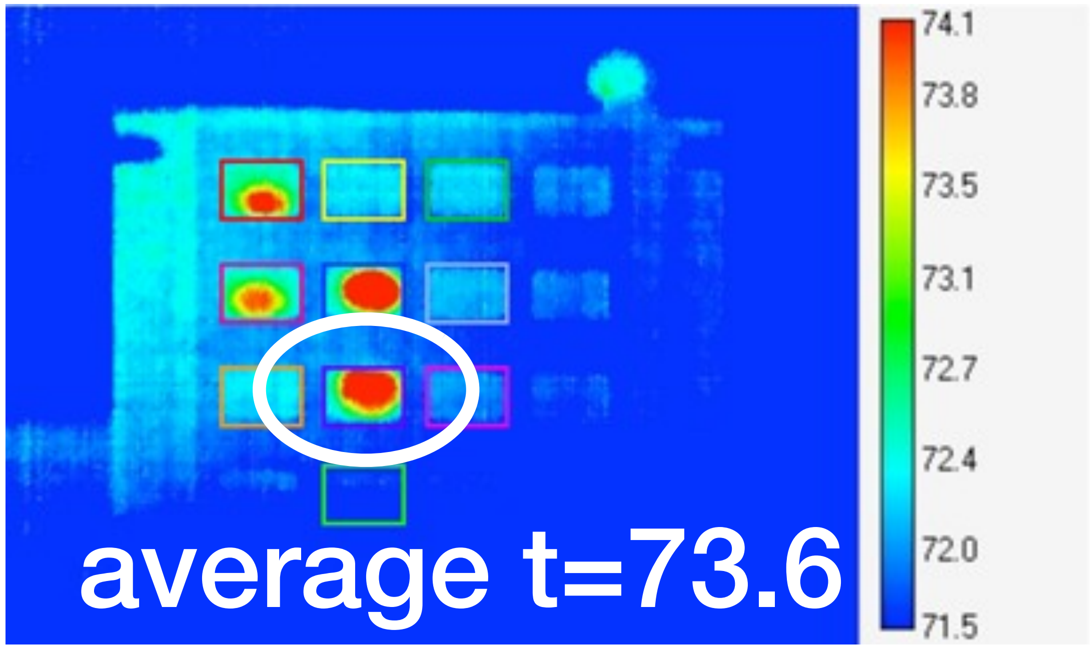
Automated review: which buttons were pressed?

Basic idea: for each region, determine if it is hot above a certain threshold



t_0=71

calibration



average t=73.6

after entry

Can repeat this process for each region, then sort in order of $\Delta = t - t_0$

Examined regions in isolation because we didn't observe much heat spread

This is the mean method, also use max and binarize variants

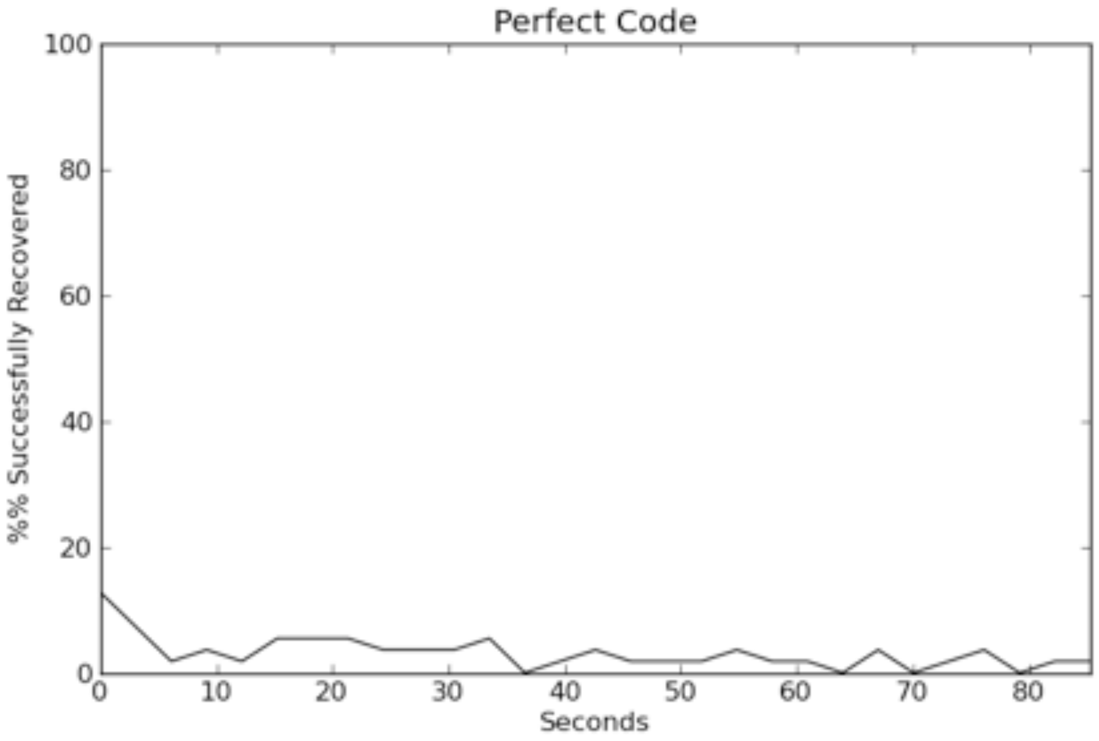
How did we do?

How did we do?

First goal: recover the exact code entered

How did we do?

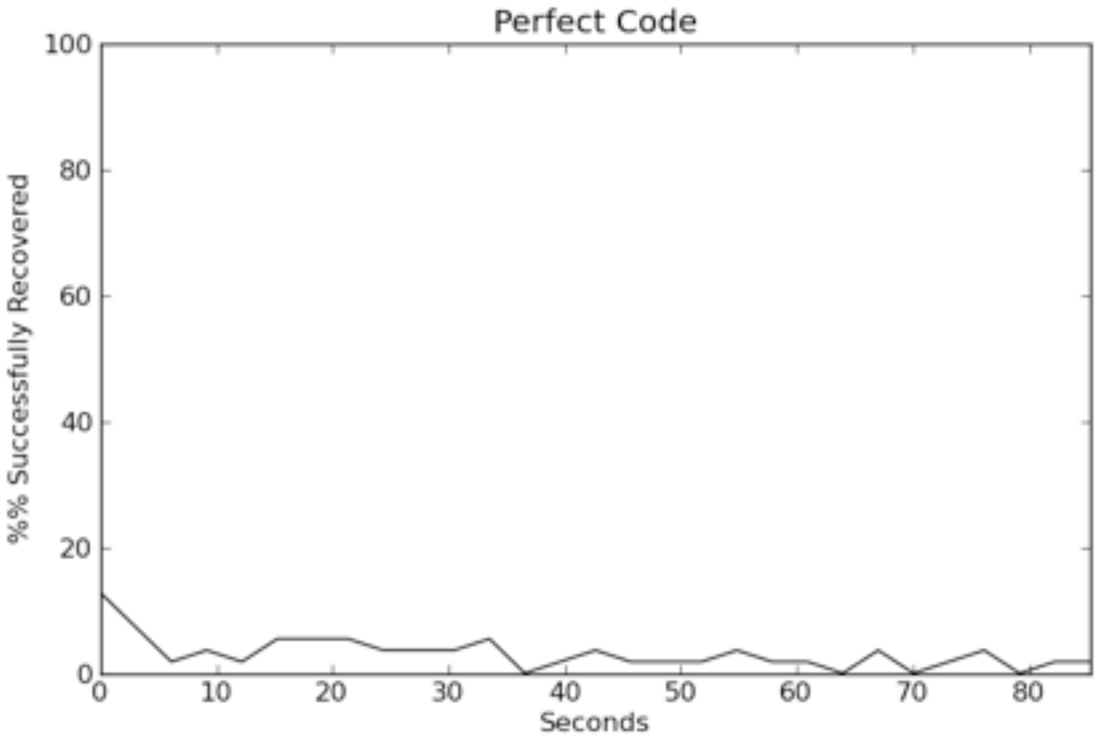
First goal: recover the exact code entered



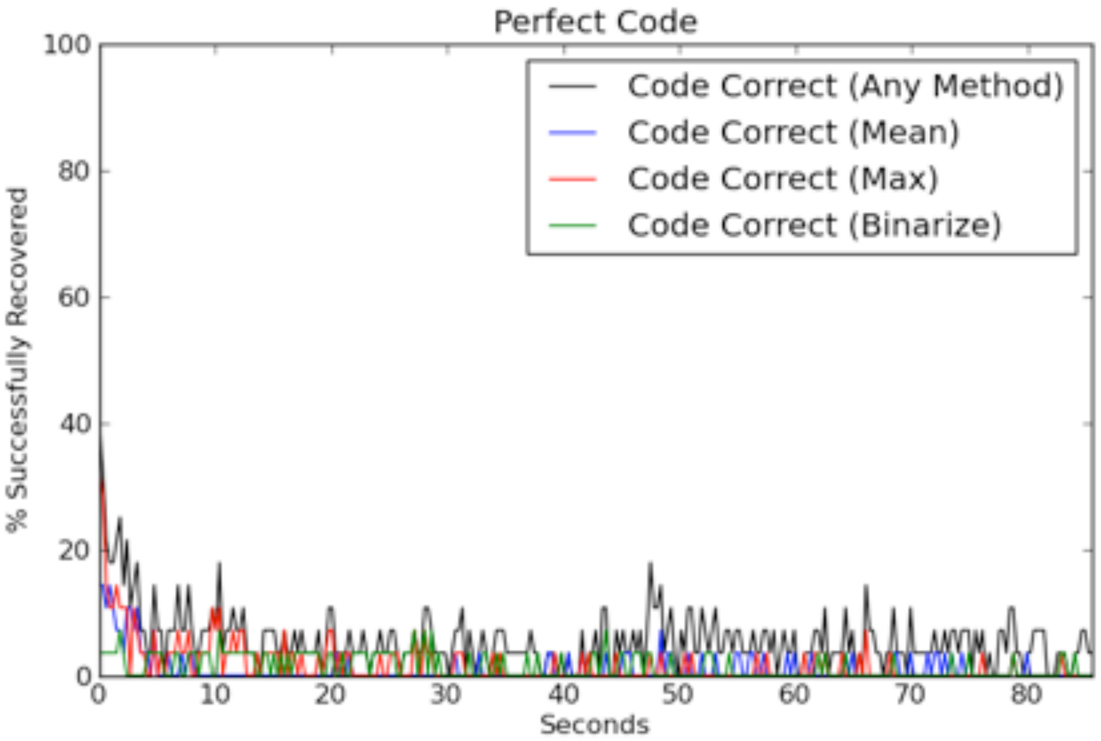
human review

How did we do?

First goal: recover the exact code entered



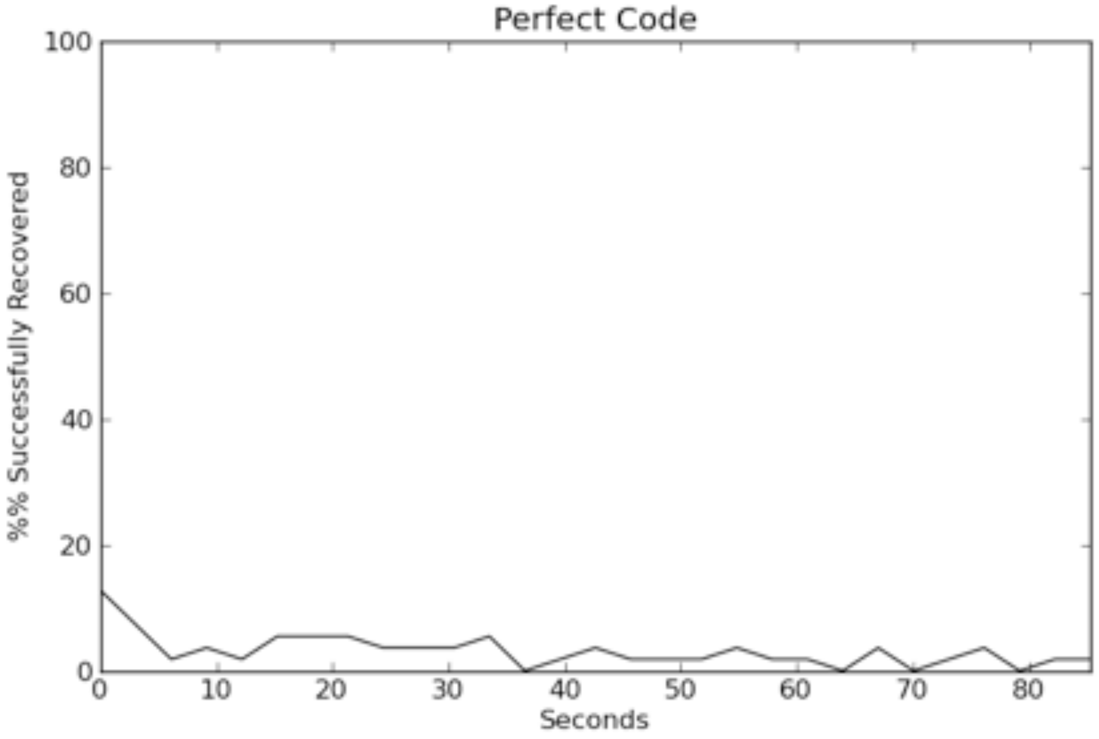
human review



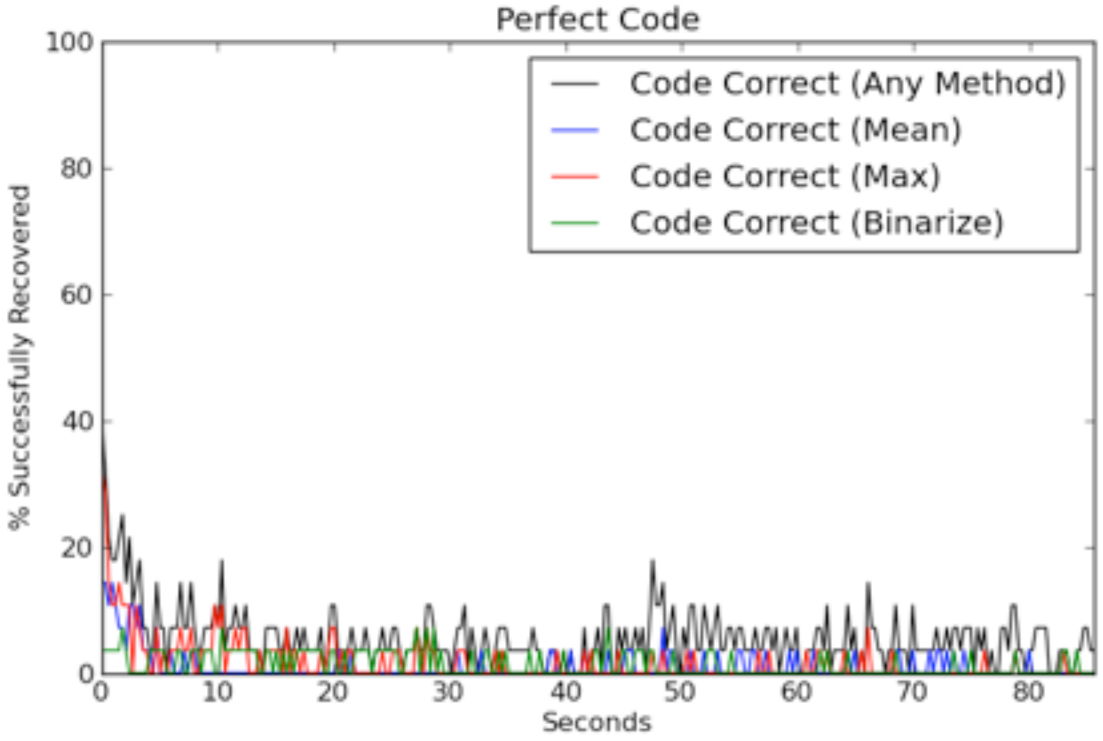
automated review

How did we do?

First goal: recover the exact code entered



human review



automated review

Bad news: the picture doesn't get much better if we allow for slight mistakes (transpositions, one wrong key, etc.)

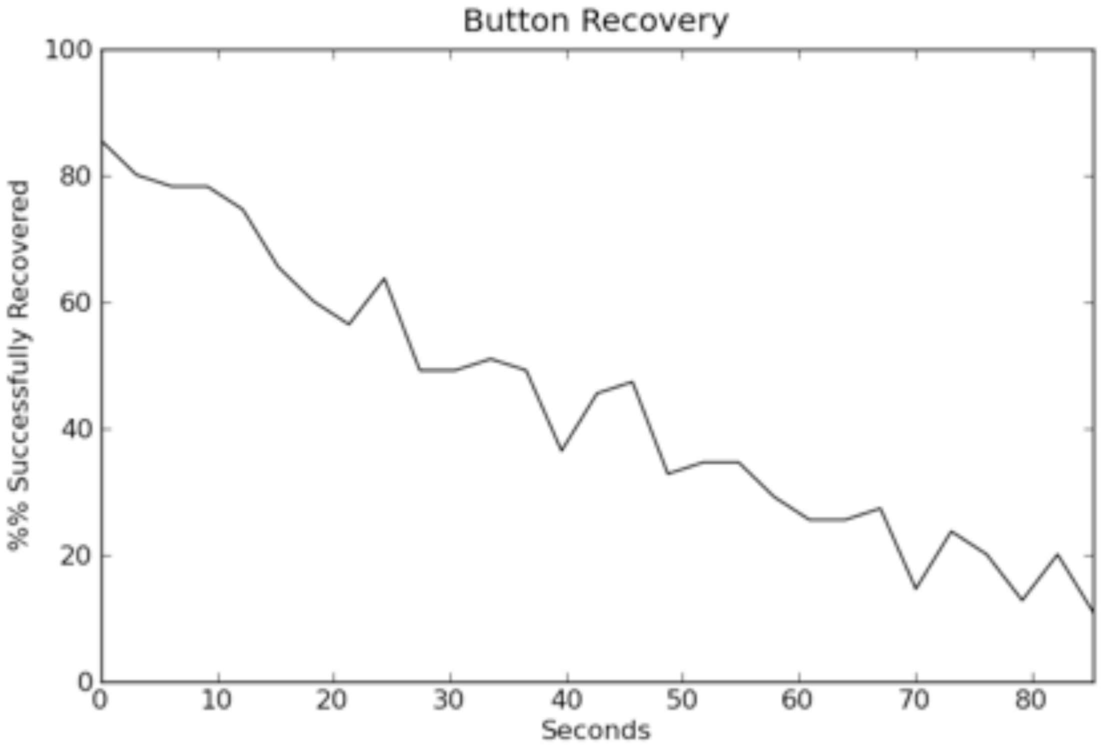
How did we do?

How did we do?

Second goal: recover the buttons pressed (not necessarily the correct order)

How did we do?

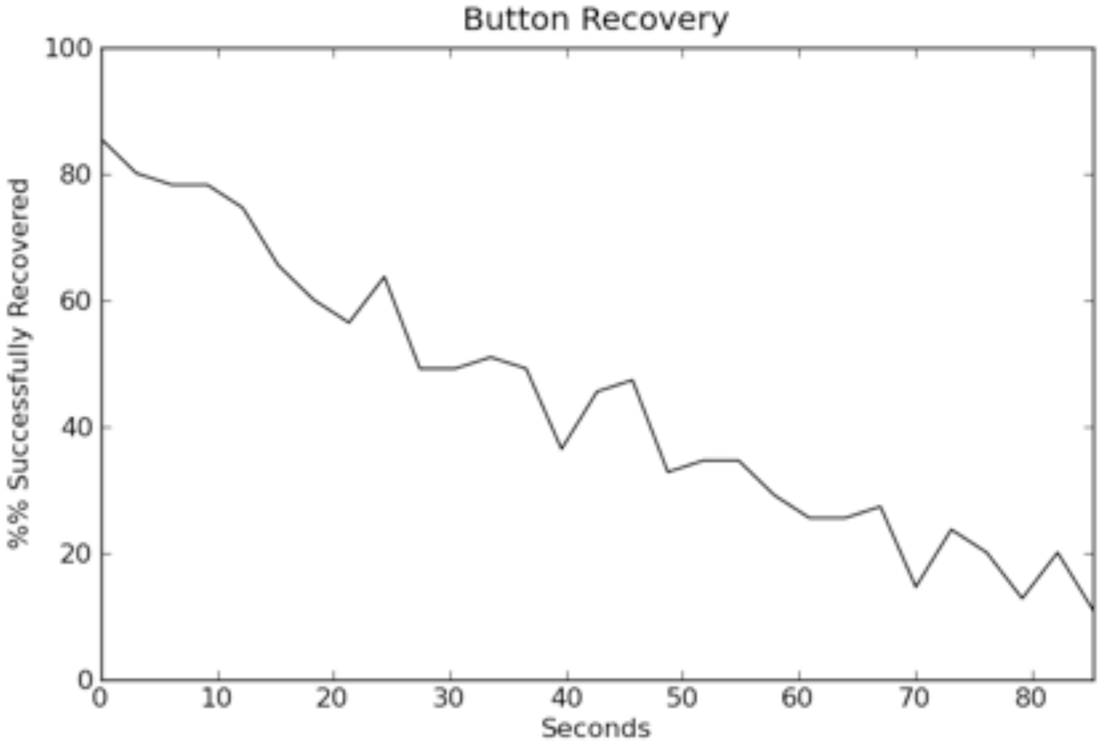
Second goal: recover the buttons pressed (not necessarily the correct order)



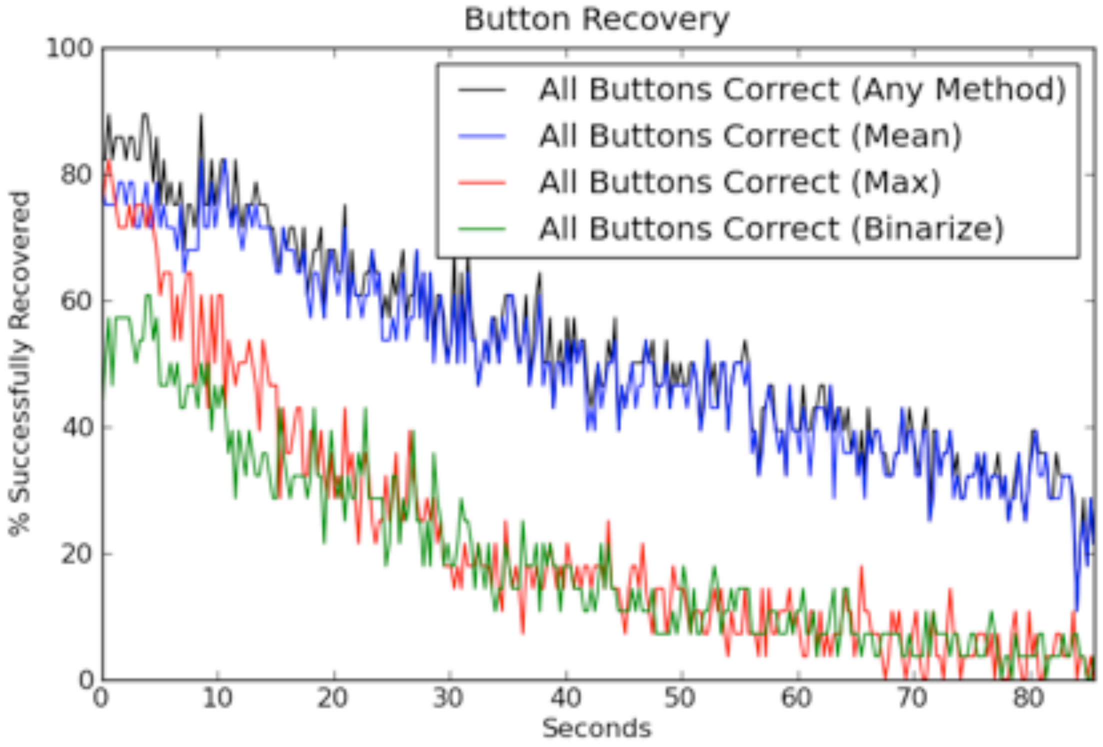
human review

How did we do?

Second goal: recover the buttons pressed (not necessarily the correct order)



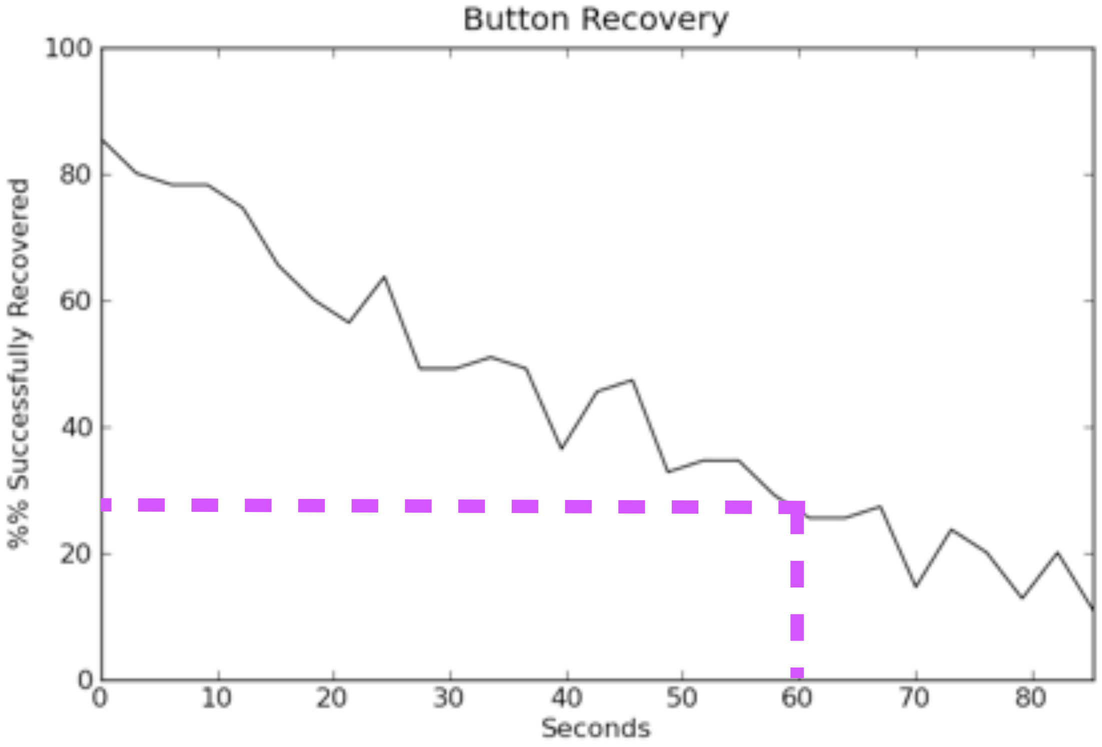
human review



automated review

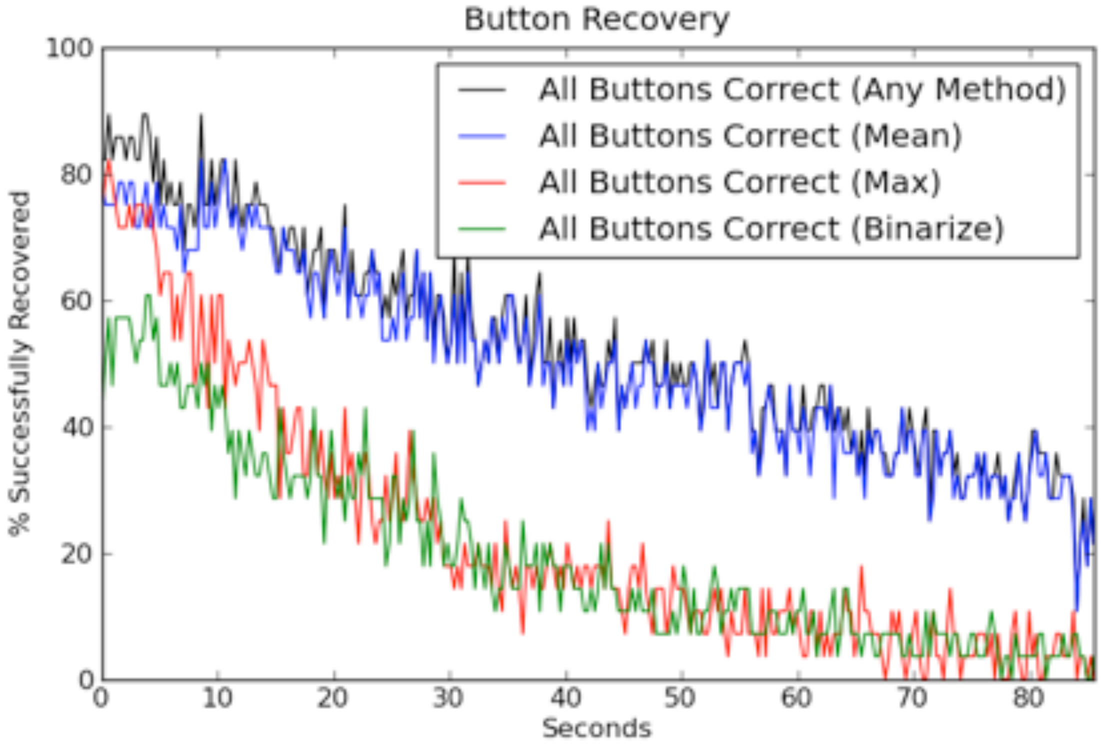
How did we do?

Second goal: recover the buttons pressed (not necessarily the correct order)



human review

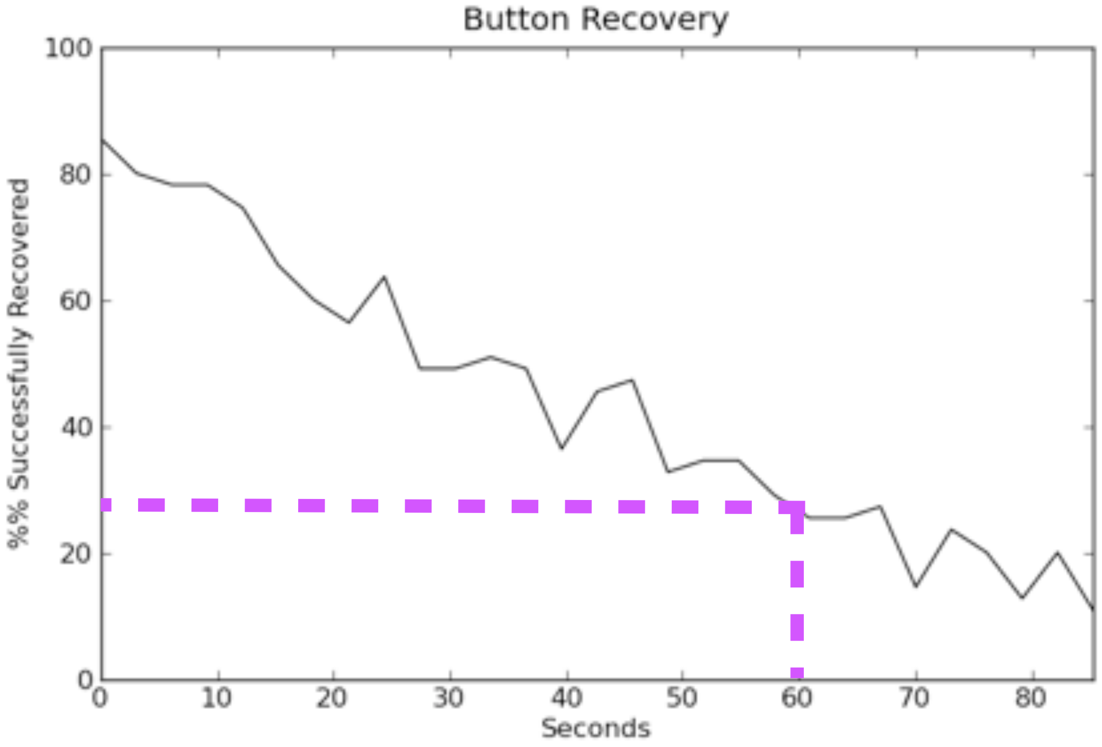
recover ~30% after 1 minute

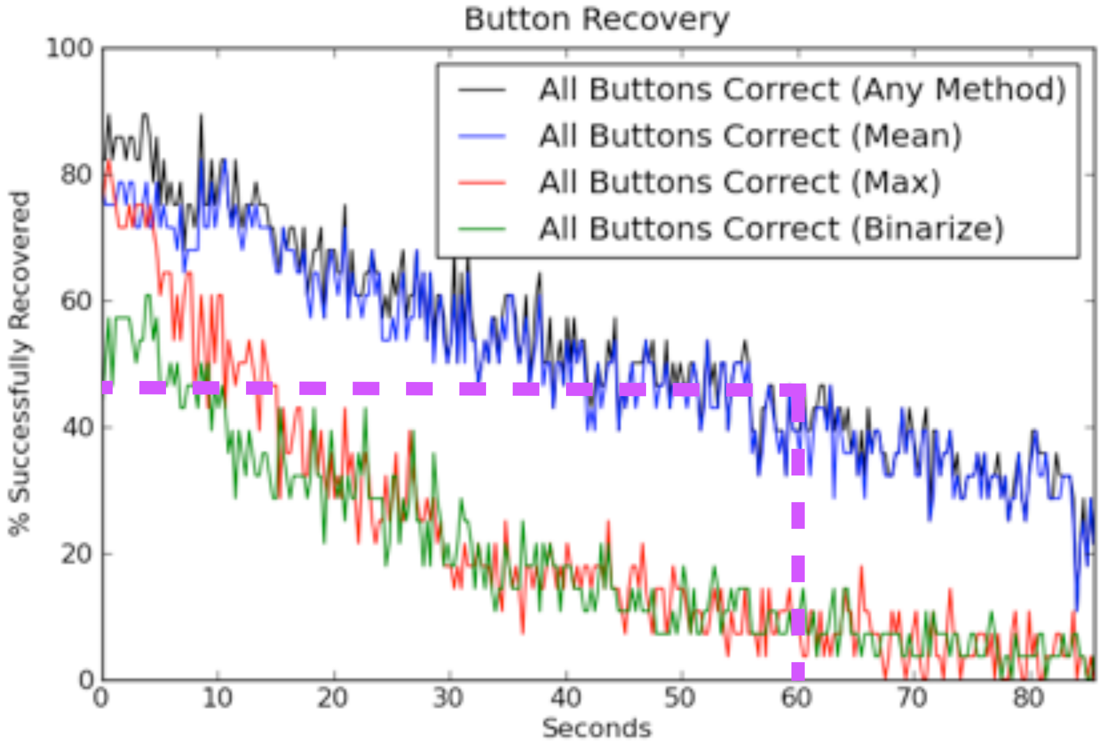


automated review

How did we do?

Second goal: recover the buttons pressed (not necessarily the correct order)





human review

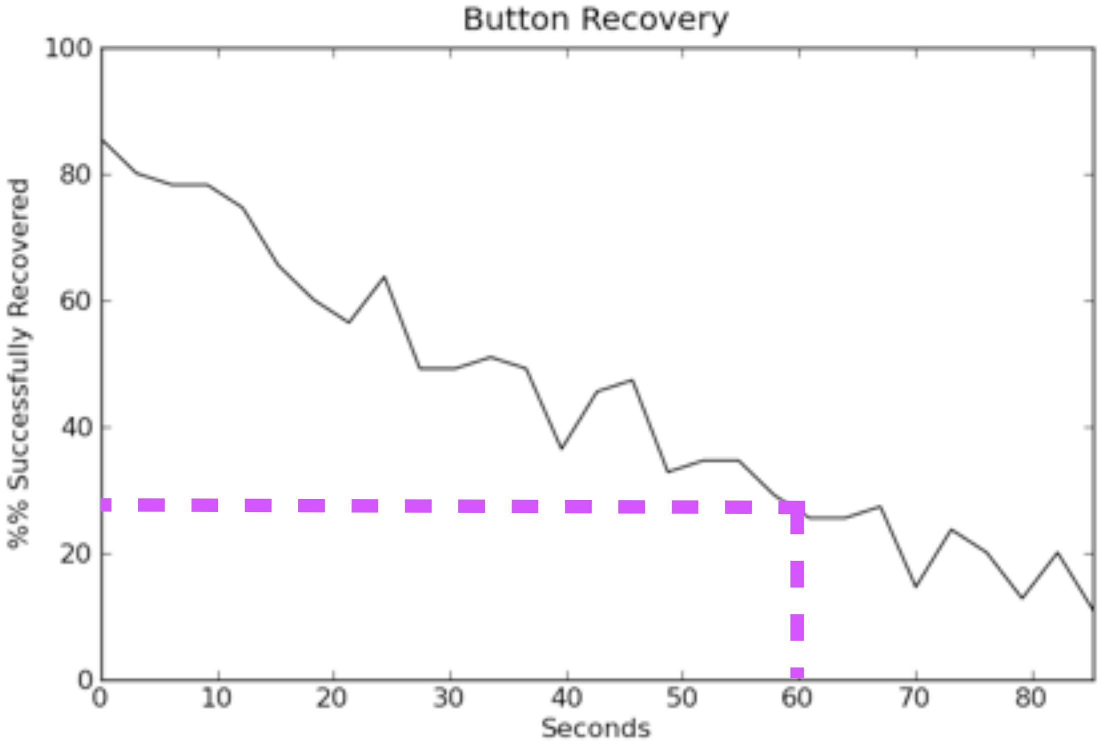
recover ~30% after 1 minute

automated review

recover ~50% after 1 minute

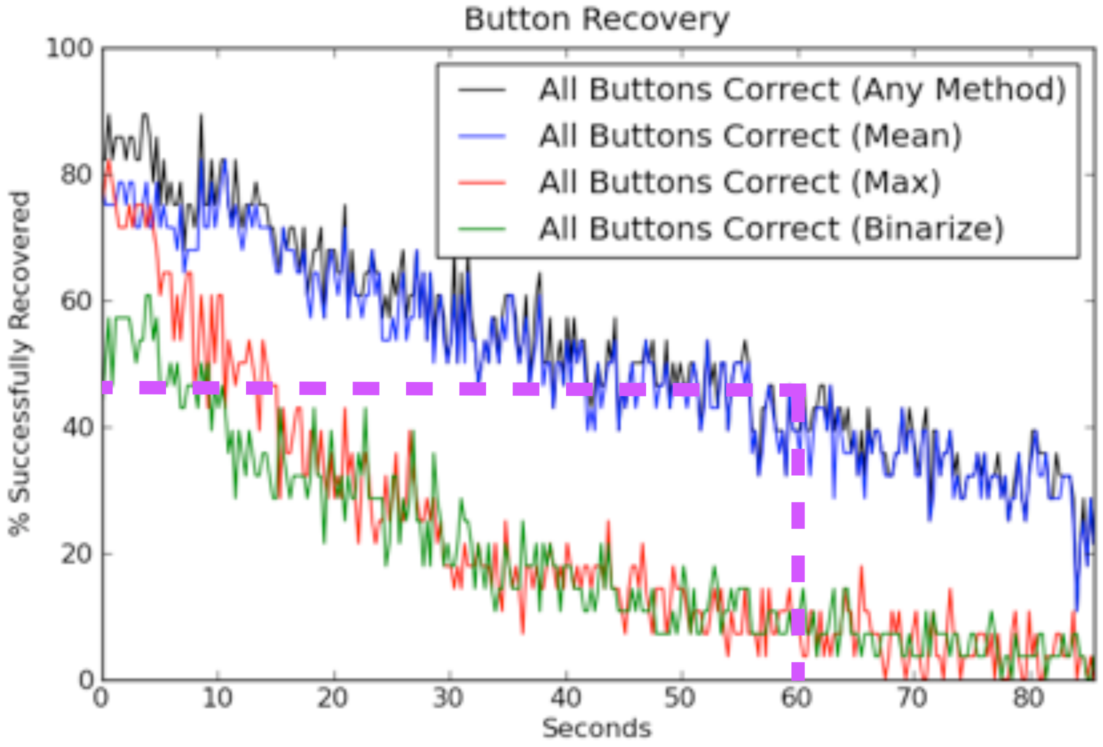
How did we do?

Second goal: recover the buttons pressed (not necessarily the correct order)



human review

recover ~30% after 1 minute



automated review

recover ~50% after 1 minute

Not only is automated review scalable, it’s also significantly more accurate

Outline

Experiment design

Camera data

Analyzing the data

Conclusions

Conclusions and future work

Conclusions and future work

Conducted study of the efficacy of thermal cameras in a variety of scenarios

Conclusions and future work

Conducted study of the efficacy of thermal cameras in a variety of scenarios

- Most effective: with plastic we recovered ~50% of codes a full minute after

Conclusions and future work

Conducted study of the efficacy of thermal cameras in a variety of scenarios

- Most effective: with plastic we recovered ~50% of codes a full minute after
- Least effective: metal keypad doesn’t work at all right now

Conclusions and future work

Conducted study of the efficacy of thermal cameras in a variety of scenarios

- Most effective: with plastic we recovered ~50% of codes a full minute after
- Least effective: metal keypad doesn’t work at all right now
- Also saw that different body temperatures and pressing styles mattered

Conclusions and future work

Conducted study of the efficacy of thermal cameras in a variety of scenarios

- Most effective: with plastic we recovered ~50% of codes a full minute after
- Least effective: metal keypad doesn't work at all right now
- Also saw that different body temperatures and pressing styles mattered

Future work and open problems:

- Use a wider set of choices: different materials, temperatures, etc.
- Analyzing footage rather than individual frames

Conclusions and future work

Conducted study of the efficacy of thermal cameras in a variety of scenarios

- Most effectiv a full minute after
- Least effectiv v
- Also saw tha tyles mattered

Thanks!
Any questions?

Future work and open problems:

- Use a wider set of choices: different materials, temperatures, etc.
- Analyzing footage rather than individual frames