Lessons in Crowdsourced Video Annotation

Carl Vondrick    Deva Ramanan
University of California, Irvine
Why is VIRAT Challenging?

Tracking Objects in MTurk

Vondrick, Ramanan, Patterson ECCV10

Annotating Events

Yuen, Russell, Liu, Torralba, ICCV 09
Why is annotation important?

Sorokin and Forsyth. CVPR 2008.

Amazon Mechanical Turk


LabelMe


PASCAL2


IMAGENET


ESP Game

Large-scale, labelled image datasets have an impact. Analogous video datasets are more scarce.
Annotation as a scientific endeavor

Surprisingly cross-disciplinary

**HCI: Interface design**

**Vision/Learning:** Ease burden on user by inserting “vision in the loop”

**Economics:** How do we motivate the “crowd” to do good work?
Annotation as a scientific endeavor

Allows us to take meaningful “baby steps” toward fully automated analysis

Visual Tracker X is better than Y because it requires less user clicks
and now... onto VIRAT
Why is VIRAT Challenging?

“lots of little cars going in many different directions. YUCK!!” — beckyp (one of our best workers)
How do we use crowdsourcing to annotate video?
Efficient Video Annotation
Efficient Video Annotation

Video Annotation with Nonlinear Motion
Video Annotation with a Visual Tracker

Learn appearance model:

\[ \phi_n(b_n) = \begin{bmatrix} HOG \\ RGB \end{bmatrix} \]

by training an SVM:

\[ w^* = \arg\min_w \frac{1}{2} w \cdot w + C \sum_{n} \max(0, 1 - y_n w \cdot \phi_n(b_n)) \]
Use constrained tracker to interpolate:

\[ b_{0:T} = \arg\min_{b_{1:T}} \sum_{t=1}^{T} \left( U_t(b_t) + S(b_t, b_{t-1}) \right) \]

\[ s.t. \quad b_0 = b_0^* \quad \text{and} \quad b_T = b_T^* \]

\[ U_t(b_t) = -w \cdot \phi_t(b_t) \]

\[ S(b_t, b_{t-1}) = \alpha \|b_t - b_{t-1}\|^2 \]

Dynamic programming + distance transforms + integral images
Video Annotation with a Visual Tracker
Video Annotation with a Visual Tracker

Vondrick, Ramanan, Patterson. ECCV 2010.
Crowdsourced Video Annotation

Split into chunks
Crowdsourced Video Annotation

Worker A
Crowdsourced Video Annotation

Worker A

Worker B
Crowdsourced Video Annotation
Crowdsourced Video Annotation

Stitch together
Quality Control

Worker A
Worker A
Quality Control

Worker A + Worker B
Quality Control

Worker A + Worker B  Bad Worker
Did this protocol work for VIRAT?
Failure Video
“Don’t waste your time with these people.” — quentin

“UGH I AM BEYOND GETTING FRUSTRATED!!!!!!” — lee45

“I guess ethical behavior is not required if you are conducting research at UCI” — avidlcfan
“I will tell everyone I know about your shoddy and unfair practices.” — Frank

Later on...

“You have apparently developed a very bad reputation with a number of the workers. Best of luck” — L.S.

**Lesson:** Good reputation attracts good workers.
“I feel strongly about my 20 cents...I expect to paid in the next 24 hours or I WILL let the IRB know ASAP.”
— quentin, student at an Ivy league university

**Lesson:** Workers feel entitled and follow principle.
“I am fine if you do not want to pay me for these, but rejecting spoils my qualifications.” — Jitendra

Lesson: Statistics are paramount to pay.
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
What went wrong?
“people seem to mark a new thing thing each frame. I sat here chewing out the computer saying ‘stay away from my car b*tch. hahahahaha.’” — lee45

**Lesson:** Failures cascade. One bad worker spoils everything.
The good, the bad, and the ugly

- **The Good**: honest workers who produce good work
- **The Bad**: honest workers who perform poorly
- **The Ugly**: dishonest workers out to scam you
How do we fix our system?
What We Did Wrong

We rejected work.
- Statistics are paramount.
- Never reject work!
- Instead, pay and block.

Workers saw previous work.
- Failures cascade.
- Workers should work independently.
- A worker claims ownership of a chunk now.

Anyone could annotate.
- Video annotation is hard and requires skill.
- Discover the good workers and only use their work.
Crowdsourced Video Annotation: Version 2.0

Split into chunks
Crowdsourced Video Annotation: Version 2.0

Worker A
Crowdsourced Video Annotation: Version 2.0

Bad Worker
Crowdsourced Video Annotation: Version 2.0

Stitch together

Worker A
Worker B
Worker C
What We Did Wrong

We rejected work. ✓
- Statistics are paramount.
- Never reject work!
- Instead, pay and block.

Workers saw previous work. ✓
- Failures cascade.
- Workers should work independently.
- A worker claims ownership of a chunk now.

Anyone could annotate.
- Video annotation is hard and requires skill.
- Discover the good workers and only use their work.
Gold Standard Video
These workers annotated all of VIRAT:

A18URWA0J27ISY  A2VB4UTFGLI6NA  A3QDBL6HYW7Y00  AXCXP0LAk7UY0  A3P03Q8LVFPJ3P  A1WP7X5H2NRFY0
A2RPQGUWZP57U  A19XW31L8EIIV6  AAKL8L5X64LWD  A3ADQ104EYHIU8  A2PF0TF53PL0JS  A1KM1SAKT19LFX
A363JB33IW99AF  A3N38YVNFQKH3V  A1937WHQ6448P  A2R6VGY0IEYI0  AEQJ5YDEP4VF5  A20IPXXUKNUN4
A198WL90Y3DQ8  AX1NALBNWBW80  A2J0HLHF7DRIYY  AU9R2KQ0KEY7M  A3R5D17QD4FIV4  A36PNAXCFAT047
A8S8E09R17AAS  A1ATGDNZ6L2PLO  AZINE6RV0JMSI  A3CD84H7AQ2RN9  A1Z3598J0371ND  A1HGKGTTTE2F2P
A2CLNXC93XH2HT  A3LMBRUSP9HWP6  A3E10ZHA0YK0ND  A34U6QLSRH29RK  ATMM8D7O3PBDU  A3LMJMCJ52FYJ9
A1QN0FA1N2PW8V  A3TD9MVDBXS3FN  A0Q06RCELZZJ  A58FN0K30XW73  A6UT4FL5GNUMZ  A1K5J0LATWC01M
A3JA2I368762HU  AK02R7J54Q131  A201JDF1C3U5SQ  A1GR7TIQT2MD1  A29MCT2JW5LIF  A2V80QG6VEECWM
A39KIZHD8UGBJD  A37Y8NP7033P1X  A1RD7BOCEWOR5T  A2HA95B0GXAT00  ANI9PXHSSHHA  A5GYHLH708ZNK
A1V52D5CN5LT2F  AD502H3OSIH77  A1Q7ROS03KLXXQ  A3TMA87177JIUJ  A1E6RS45GUAFC3  A31PSO0DCNJ4O
A18N03D7VVL96Q  A1MIMK3QQASDLW  AY85NLJN8R593  A1IU5I0C5GHY9B  A3AJG1YE503C6  A36L6LB7LC4FEV
A1QQYIGKXMA6JO  A33V10XAI3IWXS  A1LQEMQCSUPK5V  A585YACHVLZC  A305T6BFSX99Q  A1PP31MM07RE0
A2PEP36DAUK31K  A32N069DVKBTB  A16WEMVUY78TZQ  A3NA1NYJN9U2M  A26RP0Q9GJ31U  A1YAQ75LUSA02L
A34YU0YH46X6I  A2ZC16UPZXVSJ  A833VIMWCL1Q3  AQ31MQ34H6AFD  A2PJNNLH102QE9  A1RHWFPKTY1E3
ACWLRTSODR7G7  A3QHH7016UPF31  A2LSK2K7UECYCQ2  A3U0MS4W80WOB  A2GM2VIL62J38N  AH8X4M0UOVCJ
A1LXLVE8XCSBR1  A3S4CE4RJ17U13  A1D8CMLPEM60GG  AR6MC5T8YDLY  A2WJ794S0WAF3U  A2ZHS5QD168RD
A359MHHL7HGZN  A3VXXK6CMJ51G  A7K5GZTYCI7Q0  A2BERTY629PZDT  A2E41U202AYQD  A3KSSRDX4NP8N8
AALKNO596EYH  A3R0FS29N18LYT  A3EBER83NPEQY5  A223T2M1P3EFQ5  A3JEB3EY60DFQ6  A3WORR25N03L
A33STHE8LMS7N9  A19QG39H5CZCG2  A3H2SK0CIPUNP  A3TXYLD7TE3U3  A4ZL73EYDXTX  A10DHC7GKYB5Q4
A1622GWLKBQHUL  A28H662WITUB7V  A2QTVHNPBDLPR  A1U8YRLMV6J8I1  A1YQJ77SQVYVD4  AWWB5XJNC1TJ
A3DU2EWFUQCCX4  A204RAKBQ294Z  A3TL8HH9Y6CJC  A2V99E4YPE14RI  A1EX84AIPZRUX  A36541EUNCZ0WN
A3M833A9YCUCYS  A39IGUCTEQR4EK  A2059LWRZ9H545  AN9X3A3AVNZE2T  A2QTABCF83P7HS  A7HKLOUFAGRPX
A11GJ1HWB4UUGG  A2461W9Z911L1L  A3JGXVVL8147CL  ADC8VBKOVOCRR  A2KNI3S28VHTYE

137 workers
New User Interface

Annotate every object, even stationary and obstructed objects, for the entire video.

In this video, please track all of these objects:

- Car
- Person
- Bicycle
- Carried Object

Click the above button to create your first annotation.
“These are some of the best HITS I’ve ever done!”
— Joe

“UCI Computer Vision is my Favorite Requester.” – Arun

“I love these hits. They are really cool. I’m a little sad to see that there are no more” — darkangeljoy

“We are missing your hits so much waiting for you to come back.” — Umair
“I would like to tell you that your Video annotation HIT is impossible… i just wasted 5 hrs for your stupid crap” — Andrei

“It won’t even let me submit it. Says I didn’t label everything. WTF?” — dreamscaper, expert MTurk user

“I’m surprised to read so many problems. mturk is all new for me, I started 1 week ago. So far the video annotation from UCI are my favorites HITs.” — vitalogy

**Lesson:** Video annotation requires skill and patience.  
**Corollary:** Most people are bad at video annotation.
How much time do they spend?

8 months total!
Where are the workers?
How much do we pay?

5¢ per object over 10 seconds + completion bonus + Amazon fees
What’s the future of annotation technology?
Adaptive Key Frame Schedules

Not all frames are created equal.
Maximum Expected Label Change (ELC):

\[ t^* = \arg\max_{0 \leq t \leq T} \sum_{i=0}^{K} P(b_t^i) \cdot \Delta I(b_t^i) \]

\( P(b_t^i) \) is the probability worker annotates \( b_t^i \).

\( \Delta I(b_t^i) \) is the label change if \( b_t^i \) were annotated.
Adaptive Key Frame Schedules
Adaptive Key Frame Schedules

Suggests we can annotate VIRAT for **one tenth of the cost**!
Donations and a Charity Incentive

“Maybe it’s more bizarre that I keep doing these hits for a penny. I must not be the only one who finds them oddly compelling—more and more boxes show up on each hit.” — Susan

Hypothesis: Since annotation is a leisure activity, donating on behalf of the worker will encourage better annotations.
### Donations and a Charity Incentive

<table>
<thead>
<tr>
<th>Rank</th>
<th>Worker ID</th>
<th># Meals</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2EFHMHSGM9WDN</td>
<td>251.7</td>
<td>$62.92</td>
</tr>
<tr>
<td>2</td>
<td>A36541EUNCZOWN</td>
<td>245.9</td>
<td>$61.48</td>
</tr>
<tr>
<td>3</td>
<td>A7HKL0UFAGRPX</td>
<td>245.9</td>
<td>$61.48</td>
</tr>
<tr>
<td>4</td>
<td>A9LSK2K7UBYQCQ4</td>
<td>205.9</td>
<td>$51.48</td>
</tr>
<tr>
<td>5</td>
<td>A2O4RAKBQP294Z</td>
<td>192.2</td>
<td>$48.05</td>
</tr>
<tr>
<td>6</td>
<td>A5GYHLH708ZNK</td>
<td>166.0</td>
<td>$41.50</td>
</tr>
<tr>
<td>7</td>
<td>A9MK3398OP3ER</td>
<td>133.7</td>
<td>$33.43</td>
</tr>
<tr>
<td>8</td>
<td>A3ROFS92NI8LYT</td>
<td>42.1</td>
<td>$10.53</td>
</tr>
<tr>
<td>9</td>
<td>AWWW5XXJNC1TJ</td>
<td>30.4</td>
<td>$7.60</td>
</tr>
<tr>
<td>10</td>
<td>A1YQJ77SQVYYD4</td>
<td>29.7</td>
<td>$7.42</td>
</tr>
</tbody>
</table>

- **65 workers** donated **$500** to the World Food Programme.
- **2,000 cups of food.**
Thanks!

1. Lower costs with better protocols, not crowdsourcing.
2. Never reject work. But, say you do.
3. Use a crowd to discover the experts.
4. Code is available online: mit.edu/vondrick/vatic