A Mishmash of Methods for Mitigating the Model Mismatch Mess

Andrew D. Ker[a] and Tomáš Pevný[b]

[a] Department of Computer Science, Oxford University, UK.
[b] Agent Technology Center, Czech Technical University in Prague, Czech Republic.

4th February 2014

Image database to study model mismatch

- 9000 JPEG images from each of 9 Flickr users.
- All images share the same quantization table (qf = 85).
- 8 different camera models from 5 manufacturers.
- Stego images embedded with nsF5 with payload 0.05 bpp.
- 7850 dimensional *CF** features used for steganalysis.

The reality of model mismatch

Training actor \ Testing actor	1	2	3	4	5	6	7	8	9
1	0.00	0.01	0.13	0.23	0.01	0.10	0.05	0.04	0.08
2	0.02	0.00	0.16	0.10	0.02	0.10	0.04	0.04	0.06
3	0.36	0.33	0.02	0.38	0.24	0.23	0.36	0.29	0.31
4	0.03	0.01	0.18	0.00	0.01	0.09	0.03	0.04	0.06
5	0.04	0.01	0.11	0.07	0.00	0.09	0.03	0.02	0.06
6	0.06	0.02	0.16	0.04	0.03	0.06	0.07	0.03	0.07
7	0.01	0.00	0.21	0.10	0.00	0.11	0.01	0.02	0.06
8	0.05	0.03	0.12	0.09	0.01	0.09	0.04	0.01	0.08
9	0.03	0.01	0.19	0.17	0.01	0.11	0.08	0.04	0.05

Testing error of Fisher Linear Discriminant classifiers

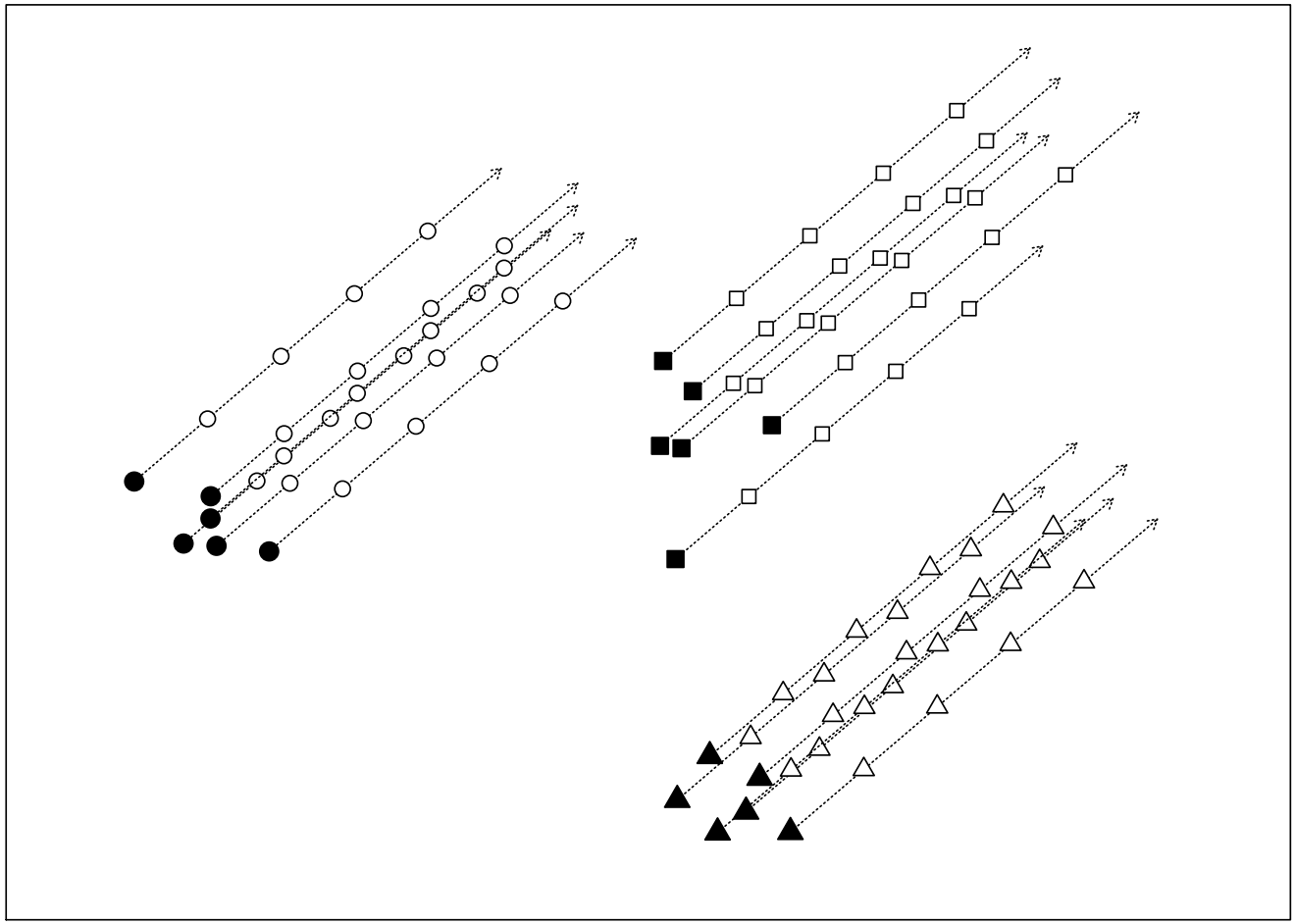
How to measure error?

- Traditional error measures do not quantify the stability of the detector.
- Mean error rate $\mu_1 = \frac{1}{k}\sum_i P_{\mathrm{E}}^i$.
- Root mean square error rate $\mu_2 = \sqrt{\frac{1}{k}\sum_i \left(P_{\mathrm{E}}^i\right)^2}$.
- Maximum error rate $\mu_\infty = \max_i P_{\mathrm{E}}^i$.

Summary of the baseline

matched cases			mismatched cases		
μ_1	μ_2	μ_∞	μ_1	μ_2	μ_∞
0.0204	0.0315	0.0663	0.0981	0.1369	0.3887

Mismatch due to shift



Mean distances of covers

...from other actor's centroid	Actor's whose centroid distance is measured...								
	1	2	3	4	5	6	7	8	9
1	0	0.59	3.93	12.58	0.66	1.90	2.82	0.72	0.79
2	1.40	0	7.27	3.67	1.35	0.38	1.07	0.48	1.02
3	12.84	12.46	0	19.02	10.56	9.25	13.92	12.00	12.58
4	5.01	0.50	8.77	0	0.70	0.37	2.01	1.97	1.06
5	0.49	1.09	2.54	5.39	0	2.17	2.29	2.56	2.00
6	2.18	0.81	5.09	1.72	0.51	0	1.04	1.09	2.08
7	0.26	0.35	7.53	2.92	0.53	0.76	0	0.49	0.10
8	1.65	1.05	2.00	5.09	1.28	1.33	1.86	0	3.69
9	0.26	0.61	4.82	3.31	0.44	0.95	1.26	1.30	0

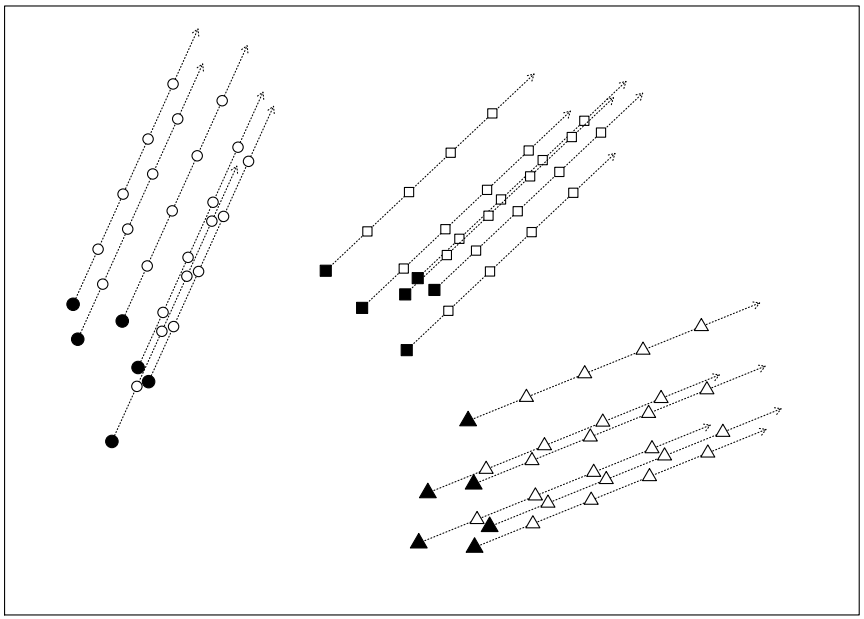
Correlation with errors: $\rho = 0.91$ $\tau = 0.51$

Centering the cover means helps

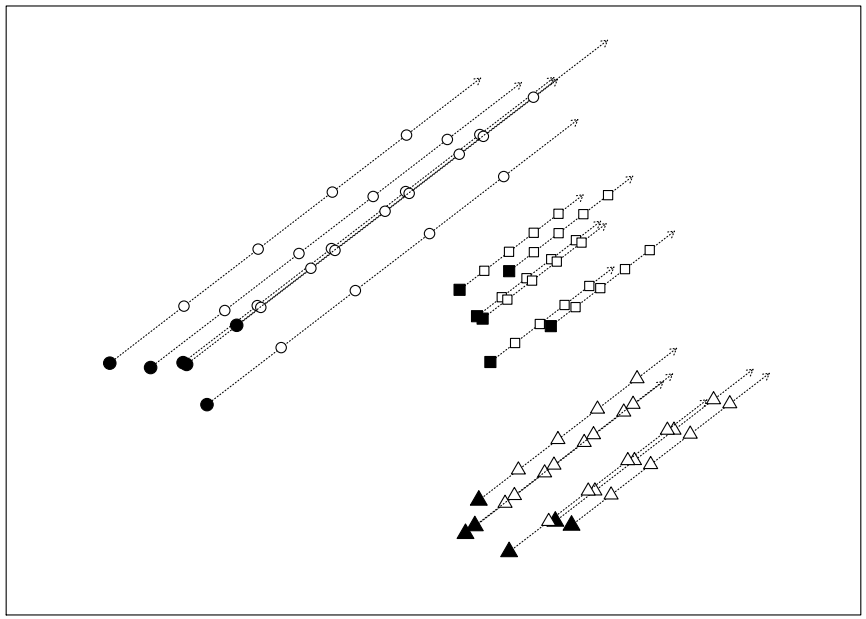
	matched cases			mismatched cases		
	μ_1	μ_2	μ_∞	μ_1	μ_2	μ_∞
Baseline	0.0204	0.0315	0.0663	0.0981	0.1369	0.3887
Subtracted mean				0.0691	0.0838	0.2332

Mismatch due to angle and speed

images moves in different angles



images moves at different speed



$\rho = -0.52,\ \tau = -0.44$

$\cos\alpha_{i,j} = \frac{w_i \cdot w_j}{\|w_i\|\|w_j\|}$

$\rho = -0.10,\ \tau = -0.09$

$r_{i,j} = \frac{\left|\|w_i\| - \|w_j\|\right|}{\sqrt{\|w_i\|\|w_j\|}}$

Calibrating for the different angle

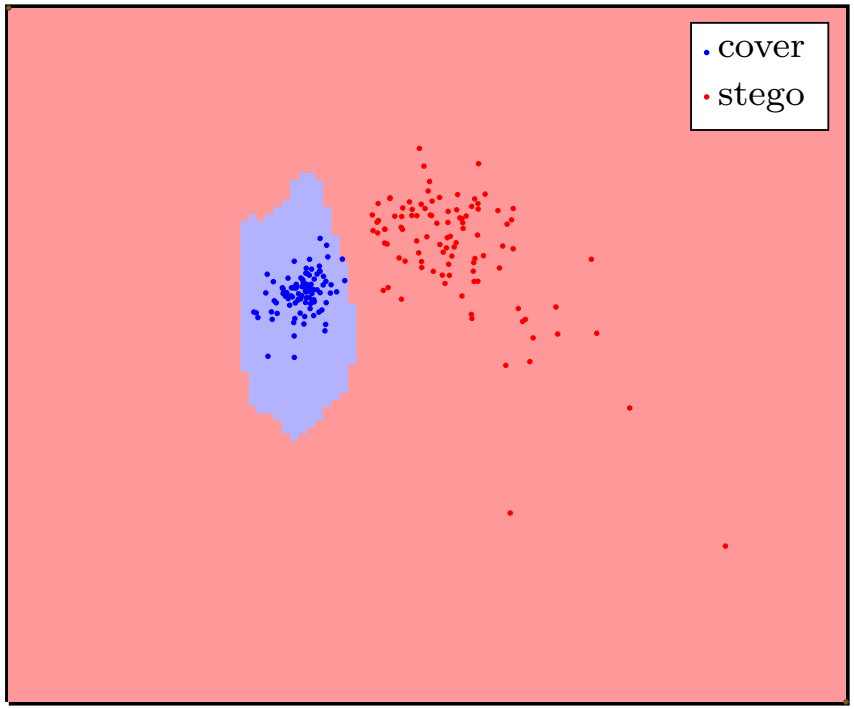
- Re-embedding of a small payload and estimating the direction of the movement
- Training an ensemble of detectors, each on a different cover source
- weighting their vote based on the alignment measured by
  - angle $\cos\alpha_i = \frac{w_i \cdot \delta}{\|w_i\|\|\delta\|}$

Experimental results

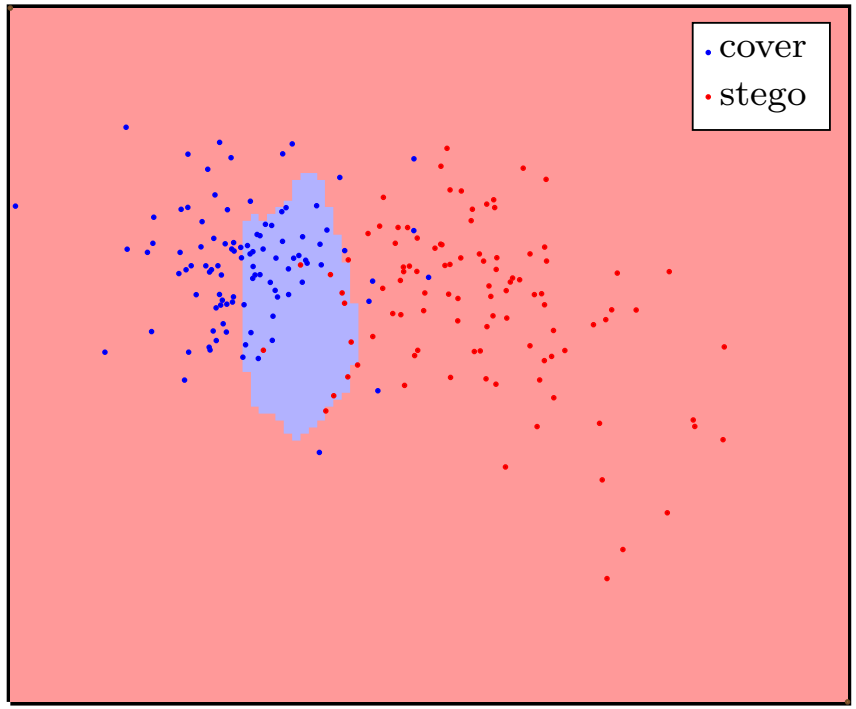
	matched cases			mismatched cases		
	μ_1	μ_2	μ_∞	μ_1	μ_2	μ_∞
Baseline	0.0204	0.0315	0.0663	0.0981	0.1369	0.3887
Centered				0.0691	0.0838	0.2332
Ensemble voting						
Equal weight				0.0439	0.0627	0.1366
Weight by $\cos\alpha_i$				0.0433	0.0593	0.1160
Ensemble voting, centered						
Equal weight				0.0391	0.0479	0.0776
Weight by $\cos\alpha_i$				0.0372	0.0468	0.0806

Mismatch due to false certainty (1)

matched case



mismatched case



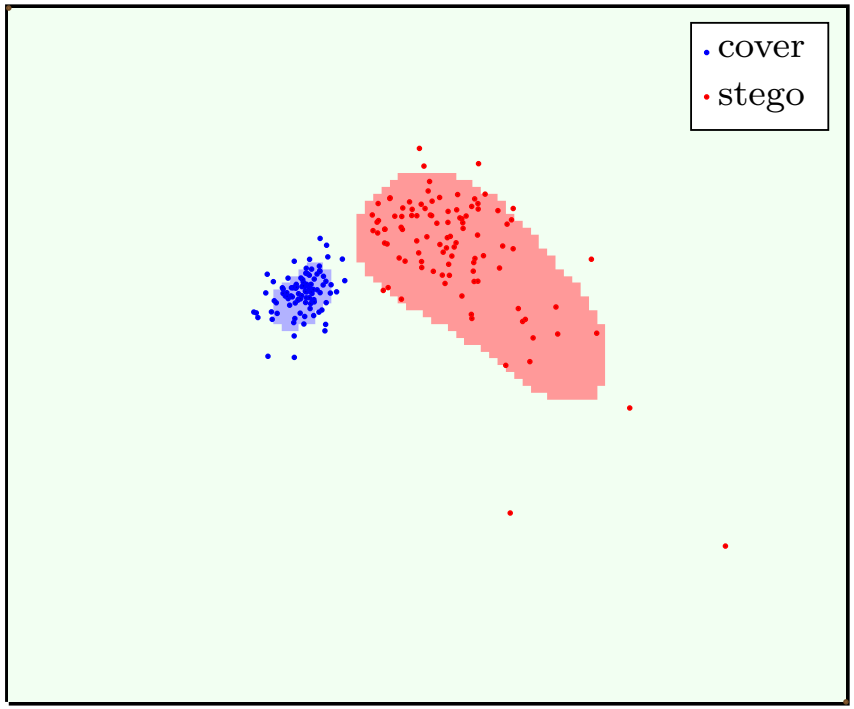
Decision areas of Binary Support Vector Machines

Abstaining classifiers

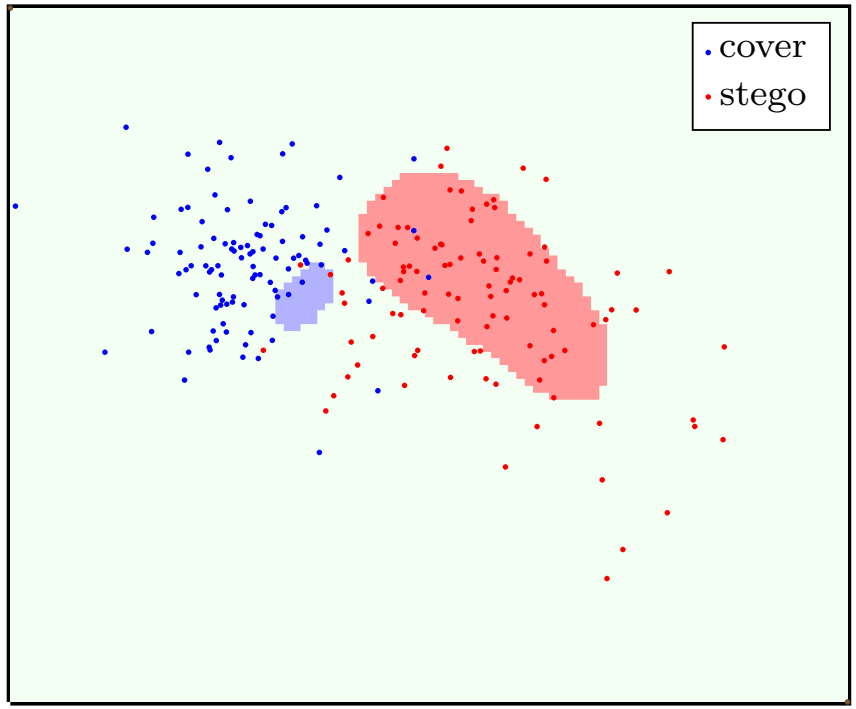
- "I don't know" is
  - honest
  - potentially advantageous in pooled steganalysis.
- Implemented as a pair of 1-class Support Vector Machines (1-SVMs).
- 1-SVMs and 2-SVMs have hyper-parameters found by cross-validation.

Mismatch due to false certainty (1)

matched case



mismatched case



Decision areas of abstaining SVMs

Measuring error of abstaining classifiers

- The value of abstaining classifiers is in the pooled steganalysis.
- Introduce deflection coefficient

$$d = \frac{(1 - P_{\mathrm{FP}} - P_{\mathrm{FN}})\sqrt{1 - P_{\mathrm{DK}}}}{\sqrt{P_{\mathrm{FP}}(1 - P_{\mathrm{FP}})} + \sqrt{P_{\mathrm{FN}}(1 - P_{\mathrm{FN}})}},$$

where P_{DK} is the probability of "don't know".

Experimental results

	mismatched cases			
	pooled error rates			d_∞
	P_{FP}	P_{FN}	P_{DK}	
2-SVMs	0.165	0.059	0.000	0.465
1-SVMs	0.028	0.010	0.539	1.017

Conclusion

- Uncorrected mismatched $\rightarrow$ error rates 5$\times$ matched case.
- Centering and matching direction $\rightarrow$ error rates 1.5-2$\times$ matched case.
- Possibility of answering "don't know".

- This is a mishmash of methods: no definitive answer.
- Research tends to be driven by the metrics:

are we chasing the right ones?