PROCEEDINGS

Electronic Imaging
Science and Technology

# *Computational Imaging VI*

**Charles A. Bouman**
**Eric L. Miller**
**Ilya Pollak**
*Editors*

**28–29 January 2008**
**San Jose, California, USA**

*Sponsored and Published by*
IS&T—The Society for Imaging Science and Technology
SPIE

*Cosponsored by*
GE Healthcare (USA)

**Volume 6814**

Proceedings of SPIE, 0277-786X, v. 6814

The papers included in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. The papers published in these proceedings reflect the work and thoughts of the authors and are published herein as submitted. The publishers are not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from this book:
Author(s), "Title of Paper," in *Computational Imaging VI*, edited by Charles A. Bouman, Eric L. Miller, Ilya Pollak, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 6814, Article CID Number (2008).

ISSN 0277-786X
ISBN 9780819469861

Copublished by
**SPIE**
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone +1 360 676 3290 (Pacific Time) · Fax +1 360 647 1445
SPIE.org
and
**IS&T—The Society for Imaging Science and Technology**
7003 Kilworth Lane, Springfield, Virginia, 22151 USA
Telephone +1 703 642 9090 (Eastern Time) · Fax +1 703 642 9094
imaging.org



Printed in the United States of America.

---

**Paper Numbering:** Proceedings of SPIE follow an e-First publication model, with papers published first online and then in print and on CD-ROM. Papers are published as they are submitted and meet publication criteria. A unique, consistent, permanent citation identifier (CID) number is assigned to each article at the time of the first publication. Utilization of CIDs allows articles to be fully citable as soon they are published online, and connects the same identifier to all online, print, and electronic versions of the publication. SPIE uses a six-digit CID article numbering system in which:

- The first four digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B … 0Z, followed by 10-1Z, 20-2Z, etc.

The CID number appears on each page of the manuscript. The complete citation is used on the first page, and an abbreviated version on subsequent pages. Numbers in the index correspond to the last two digits of the six-digit CID number.

# Contents

**IMAGE ANALYSIS II**



**POSTER SESSION**

# Conference Committee

*Symposium Chair*

**Nitin Sampat,** Rochester Institute of Technology (USA)

*Conference Chairs*

**Charles A. Bouman,** Purdue University (USA)
**Eric L. Miller,** Tufts University (USA)
**Ilya Pollak,** Purdue University (USA)

*Program Committee*

**Samit Basu,** GE Global Research (USA)
**Thomas S. Denney,** Auburn University (USA)
**Peter C. Doerschuk,** Purdue University (USA)
**Peyman Milanfar,** University of California, Santa Cruz (USA)
**Joseph A. O'Sullivan,** Washington University in St. Louis (USA)
**Zygmunt Pizlo,** Purdue University (USA)
**Stanley J. Reeves,** Auburn University (USA)
**Yongyi Yang,** Illinois Institute of Technology (USA)

*Session Chairs*

Keynote Presentation
**Charles A. Bouman,** Purdue University (USA)

Image Reconstruction I
**Eric L. Miller,** Tufts University (USA)

Geometry-based Techniques in Image Analysis
**Mary L. Comer,** Purdue University (USA)

Segmentation
**Charles A. Bouman,** Purdue University (USA)

Sparse Recovery and Compressed Sensing
**Ilya Pollak,** Purdue University (USA)

Image Analysis I
**Paul Salama,** Indiana University-Purdue University at Indianapolis (USA)

Image Reconstruction II
**Edward J. Delp,** Purdue University (USA)

Image Analysis II
**Mireille Boutin,** Purdue University (USA)

## Stationary Features and Cat Detection

Donald Geman

joint work with

Francois Fleuret

Computational Imaging VI

January 29, 2008

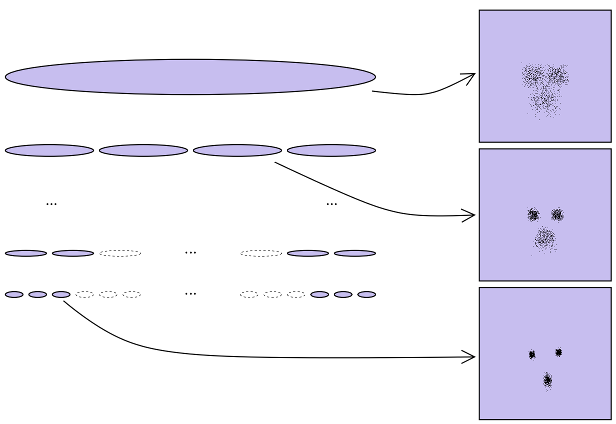
## Introduction

### Detection



## Introduction

### Detection (cont.)





## Hierarchy of Classifiers

**Advantages**

- Highly efficient scene parsing;
- Organized focusing on hard examples.

**Disadvantages**

- Many classifiers to learn;
- Inefficient learning (unless training examples can be synthesized).

x

## Introduction
### Related work

**Part-based Models**

- Constellation (M. Weber, M. Welling, P. Perona),
- Composite (D. J. Crandall, D. P. Huttenlocher),
- Implicit Shape (E. Seemann, M. Fritz, B. Schiele),
- Patchwork of Parts (Y. Amit, A. Trouvé),
- Compositional (S. Geman).

**Wholistic**

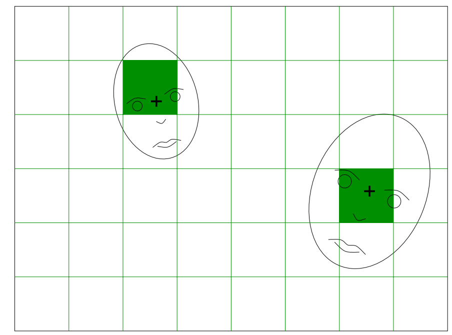
- Convolutional Networks (Y. Lecun),
- Boosted Cascades (P. Viola, M. Jones),
- Bag-of-Features (A.Zisserman, F.F.Li,).



## Introduction
### Pose space

Let $\mathcal{Y}$ be the space of poses and $\mathcal{Y}_1, \dots, \mathcal{Y}_K$ a partition.

Let $I$ be an image and $\mathbf{Y} = (Y_1, \dots, Y_K)$, where, for each $k$, $Y_k$ is a Boolean variable stating if there is an object in $I$ with pose in $\mathcal{Y}_k$.

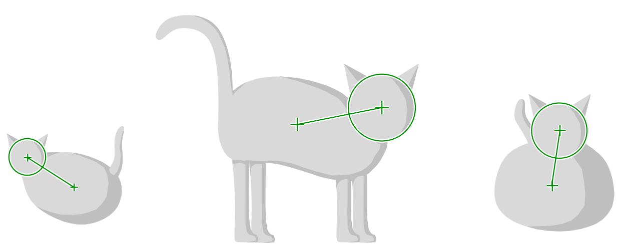




## Introduction
### Pose space (cont.)

For faces, typically $y = (u_c, v_c, \theta, s)$. Cats are more complex.

A coarse description is $y = (u_h, v_h, s_h, u_b, v_b)$.





## Training a detector
### Fragmentation

Given a training set

$$\left\{ \left( I^{(t)}, \mathbf{Y}^{(t)} \right) \right\}_t ,$$

we are interested in building, for each $k$, a classifier

$$f_k : \mathcal{I} \rightarrow \{0, 1\}$$

for predicting $Y_k$.

Without additional knowledge about the relationship between $k$, $Y_k$ and $I$, we would train $f_k$ with

$$\left\{ \left( I^{(t)}, Y_k^{(t)} \right) \right\}_t .$$

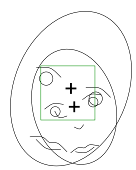
Hence, each $f_k$ is trained with a fragment ($\simeq 1/K$) of the positive population.

## Training a detector
### Aggregation

To avoid fragmentation, samples are often normalized in pose. Let

$$\xi : \mathcal{I} \to \mathbb{R}^N$$

denote an $N$-dimensional vector of features. Let $\psi : \{1, \dots, K\} \times \mathcal{I} \to \mathcal{I}$ be a transformation such that the conditional distribution of $\xi(\psi(k, I))$ given $Y_k = 1$ does not depend on $k$.





## Training a detector
### Aggregation (cont.)

Then train a single classifier $g$ with the training set

$$\left\{ \left( \psi\left(k, I^{(t)}\right),\ Y_k^{(t)} \right) \right\}_{t,k}$$

and define

$$f_k(I) = g(\psi(k, I)).$$

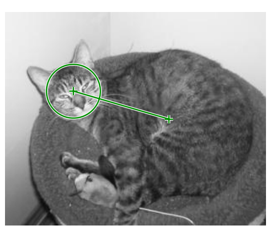
Here, samples are aggregated for training: all positive samples are used to build each $f_k$.

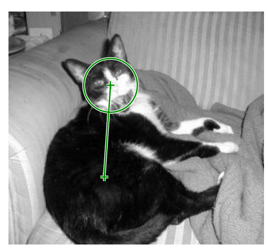


## Training a detector
### Aggregation (cont.)

However:

- Evaluating $\psi$ is computationally intensive for any non-trivial transformation.
- The mapping $\psi$ does not exist for a complex pose.





Hence, in practice, fragmentation is still the norm to deal with many deformations. But how does on deal with complex poses?



## Stationary features
### Introduction

We directly index the features with the pose.

$$\begin{tikzcd} \{1, \dots, K\} \times \mathcal{I} \arrow[r, "\psi"] \arrow[dr, "\mathbf{x}"'] & \mathcal{I} \arrow[d, "\xi"] \\ & \mathbb{R}^N \end{tikzcd}$$

## Stationary features
### Definition

Hence, we define a family of pose-indexed features as a mapping

$$\mathbf{X} : \{1, \ldots, K\} \times \mathcal{I} \to \mathbb{R}^N$$

such that they are stationary in the following sense: for every $k \in \{1, \ldots, K\}$, the probability distribution

$$P(\mathbf{X}(k, I) = x \mid Y_k = 1), \;\; x \in \mathbb{R}^N$$

does not depend on $k$.



## Stationary features
### Interpretation

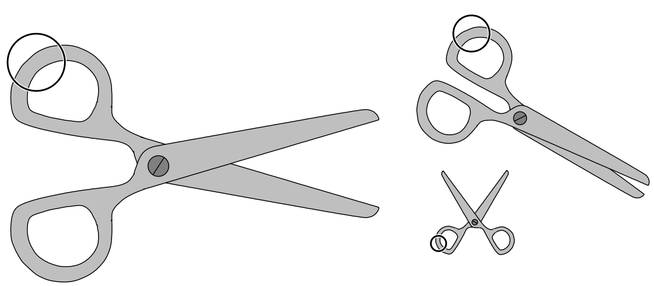
Suppose there is exactly one object in $I$ ($\sum_k Y_k = 1$) and let $Z = Z(\mathbf{Y}) \in \{1, ..., K\}$ be the corresponding pose cell.
Then stationarity means that

$$\mathbf{X}(Z) \perp Z$$

In other words, knowing $\mathbf{X}(Z)$, the feature values, tells you nothing about $Z$, the pose.



## Stationary features
### Definition (cont.)

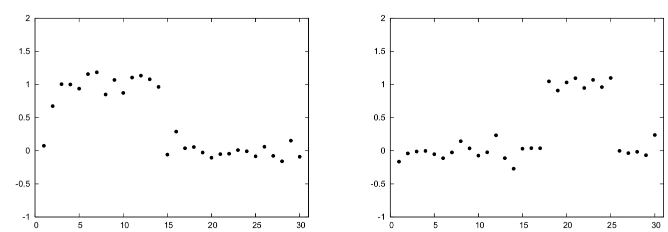




## Stationary features
### Toy example, 1D signal

$$\mathcal{Y} = \left\{(\theta_1, \theta_2) \in \{1, \ldots, N\}^2,\, 1 < \theta_1 < \theta_2 < N\right\}$$

$$P(I \mid Y = (\theta_1, \theta_2)) \;=\; \prod_{n<\theta_1} \phi_0(I_n) \prod_{\theta_1 \leq n \leq \theta_2} \phi_1(I_n) \prod_{\theta_2 < n} \phi_0(I_n)$$



$$\mathbf{X}((\theta_1, \theta_2), I) = (I(\theta_1 - 1), I(\theta_1), I(\theta_2), I(\theta_2 + 1))$$

$$P(\mathbf{X} = (x_1, x_2, x_3, x_4) \mid Y = (\theta_1, \theta_2)) = \phi_0(x_1)\phi_1(x_2)\phi_1(x_3)\phi_0(x_4)$$

## Stationary features
### Training

We then define

$$f_k(I) = g(\mathbf{X}(k, I)), \;\; k \in \{1, \dots, K\},$$

where one single classifier $g\colon \mathbb{R}^N \to \{0,1\}$ is trained with

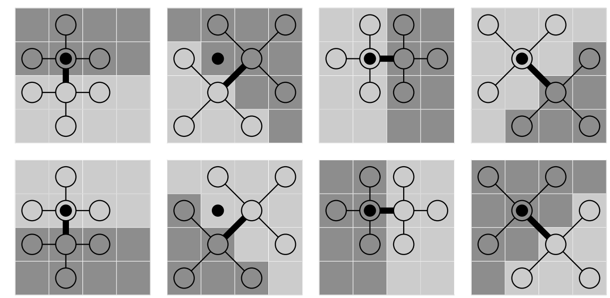
$$\left\{\left(\mathbf{X}\left(k, I^{(t)}\right),\; Y_k^{(t)}\right)\right\}_{t,k}.$$

Notice that stationarity is the main condition required to ensure that the training samples are identically distributed.



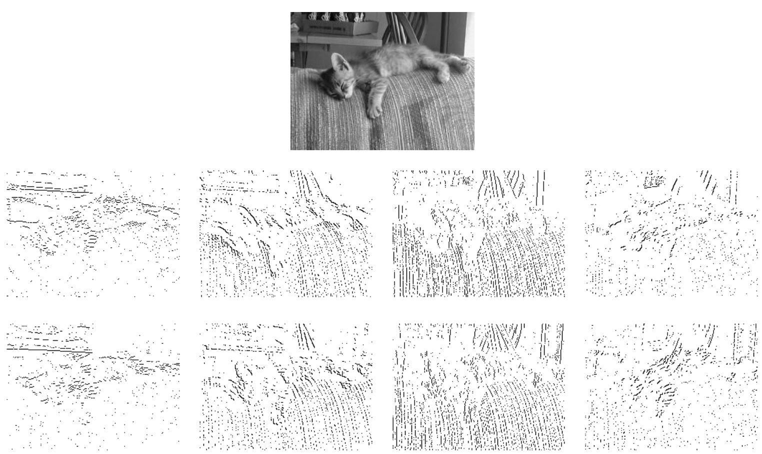
## Cat detection
### Edge detectors

Let $e_{\phi,\sigma}(I, u, v)$ be the presence on an edge in image $I$ at location $(u, v)$ with orientation $\phi \in \{0, \dots, 7\}$ at scale $\sigma \in \{1, 2, 4\}$.





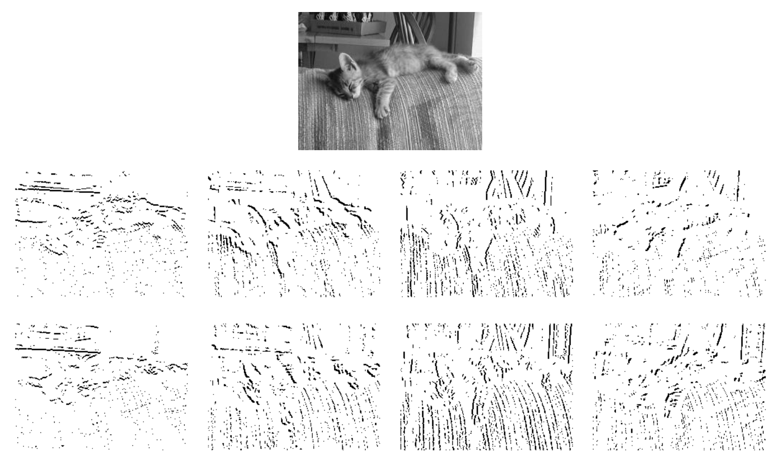
## Cat detection
### Edge detectors (cont.)



$\sigma = 1$

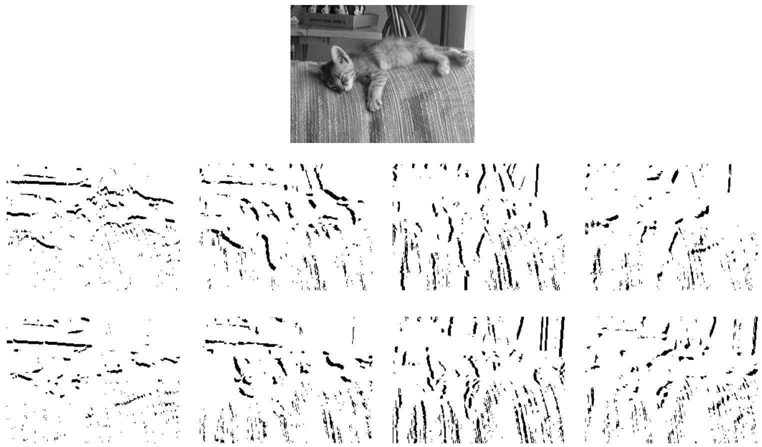


## Cat detection
### Edge detectors (cont.)



$\sigma = 2$

## Cat detection
### Edge detectors (cont.)



$\sigma = 4$



## Cat detection
### Base features

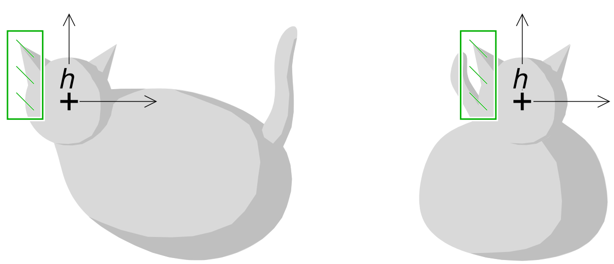
We use three types of stationary features:

- Proportion of an edge $(\phi, \sigma)$ in a window $W$.
- $L^1$-distance between the histograms of orientations at scale $\sigma$ in two windows $W$ and $W'$.
- $L^1$-distance between the histograms of gray-levels in two windows $W$ and $W'$.

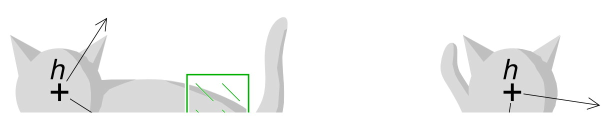


## Cat detection
### Stationary features

The windows $W$ and $W'$ are indexed either with respect to the head or with respect to the head-belly.




Head registration






## Cat detection
### Classifier

We build classifiers with Adaboost and an asymmetric weighting by sampling. If the weighted error rate is

$$L(h) = \sum_{t,k} \omega_{t,k} \mathbb{1}_{\left\{ h\left(k, I^{(t)}\right) \neq Z_k^{(t)} \right\}},$$

we use all the positive samples, and sub-sample negative ones with

$$\mu(k,t) \propto \omega_{t,k} \mathbb{1}_{\left\{ Y_k^{(t)} = 0 \right\}}.$$

Total of 2327 scenes containing 1683 cats, 85% for training.

## Folded Hierarchies
### Summary

Experiments (not shown) demonstrate:

- Fragmentation applied to complex poses is disastrous in practice;
- Naive, brute-force exploration of the pose space is computationally intractable.

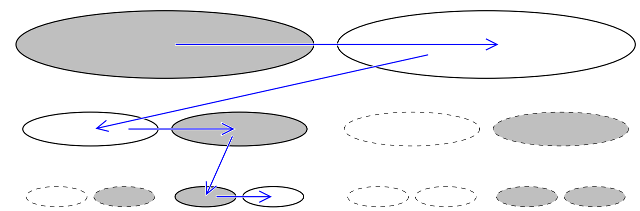
Alternatively:

- Stationary features largely avoid fragmentation;
- Hierarchical search concentrates computation on ambiguous regions.

We call this alternative a *folded hierarchy of classifiers*.

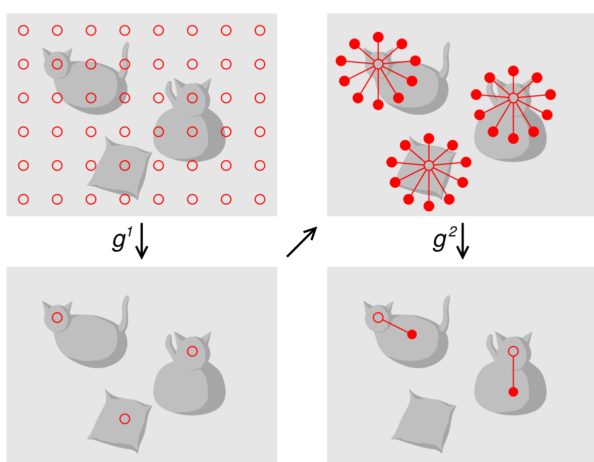


## Folded Hierarchies
### Summary (cont.)



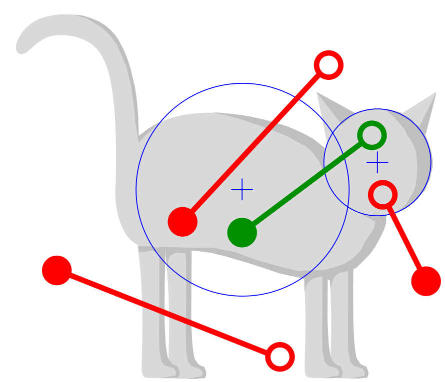


## Scene Parsing
### Strategy






## Scene Parsing
### Error criterion



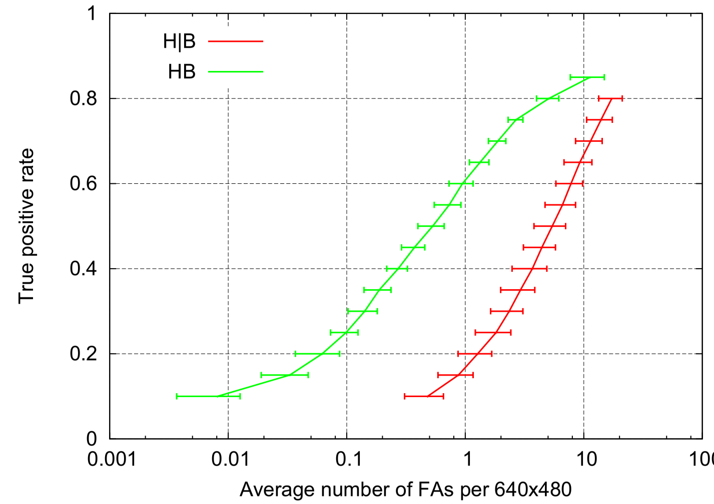
## SCENE PARSING
### ERROR RATES

| TP | H\|B | HB |
|---|---|---|
| 85% | – | 11.29 (3.61) |
| 80% | 17.23 (3.87) | 5.07 (1.08) |
| 70% | 11.40 (2.89) | 1.88 (0.32) |
| 60% | 7.80 (1.98) | 0.95 (0.22) |
| 50% | 5.41 (1.62) | 0.53 (0.13) |

Average number of FAs per $640 \times 480$

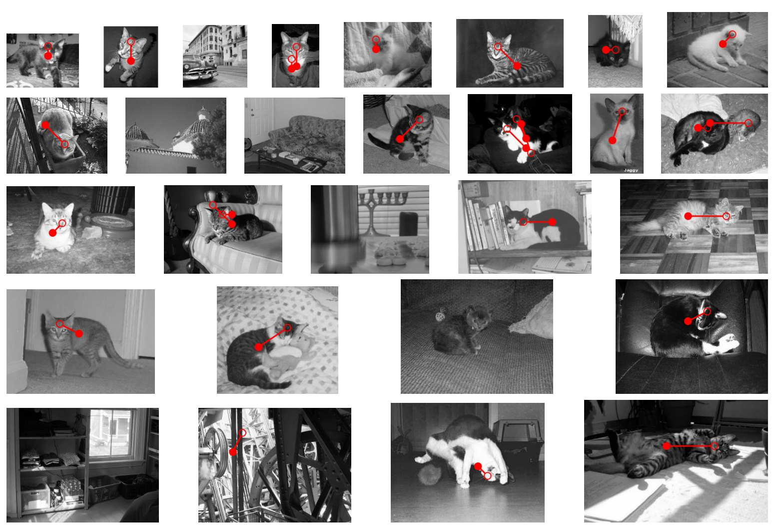


## SCENE PARSING
### ERROR RATES (CONT.)



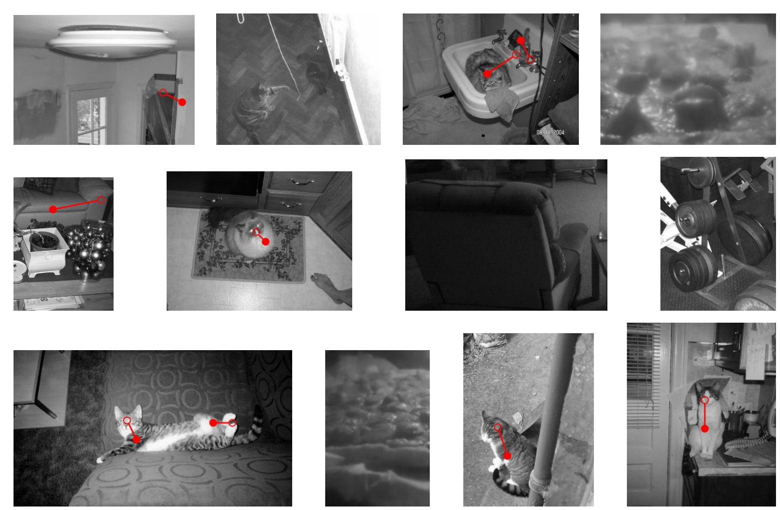



## SCENE PARSING
### RESULTS (PICKED AT RANDOM)

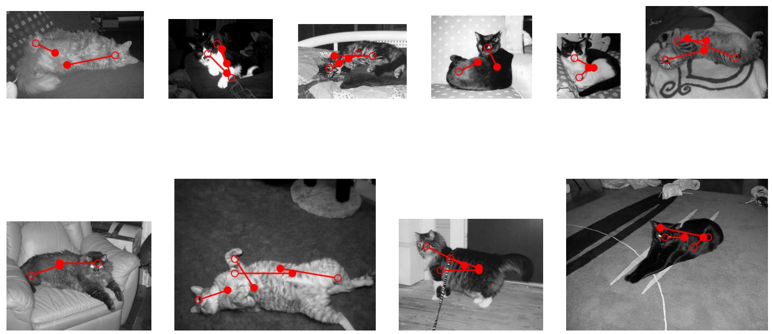




## SCENE PARSING
### RESULTS (PICKED AT RANDOM, CONT.)

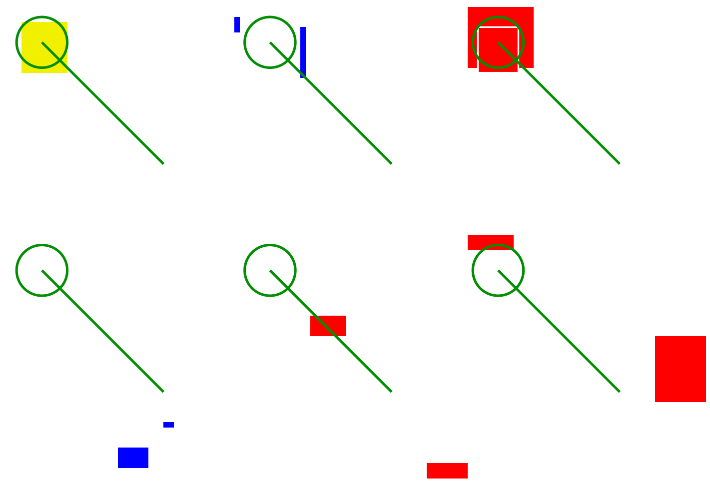


## Scene parsing
### Results (selected false alarms)





## Scene parsing
### Selected stationary features





## Conclusion

A folded hierarchy of classifiers is highly efficient:

- For (offline) learning;
- For (online) scene processing.

It combines the strengths of:

- Template matching;
- Powerful, wholistic machine learning;
- Hierarchical search.

However, this is achieved at the expense of

- Rich annotation of the training samples.
- Designing stationary features.



## Acknowledgment