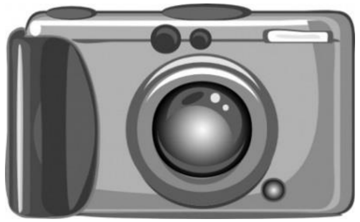
Reconstruction-free Inference on Compressive Measurements

Suhas Lohit, Kuldeep Kulkarni, Pavan Turaga, Jian Wang, Aswin Sankaranarayanan
Arizona State University Carnegie Mellon University

Inference in the Traditional Sensing Framework

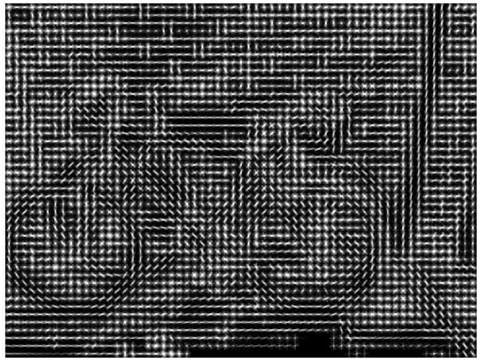
Traditional computer vision pipeline – Oracle Sensing



Camera with a
CCD/CMOS array

Feature
Extraction

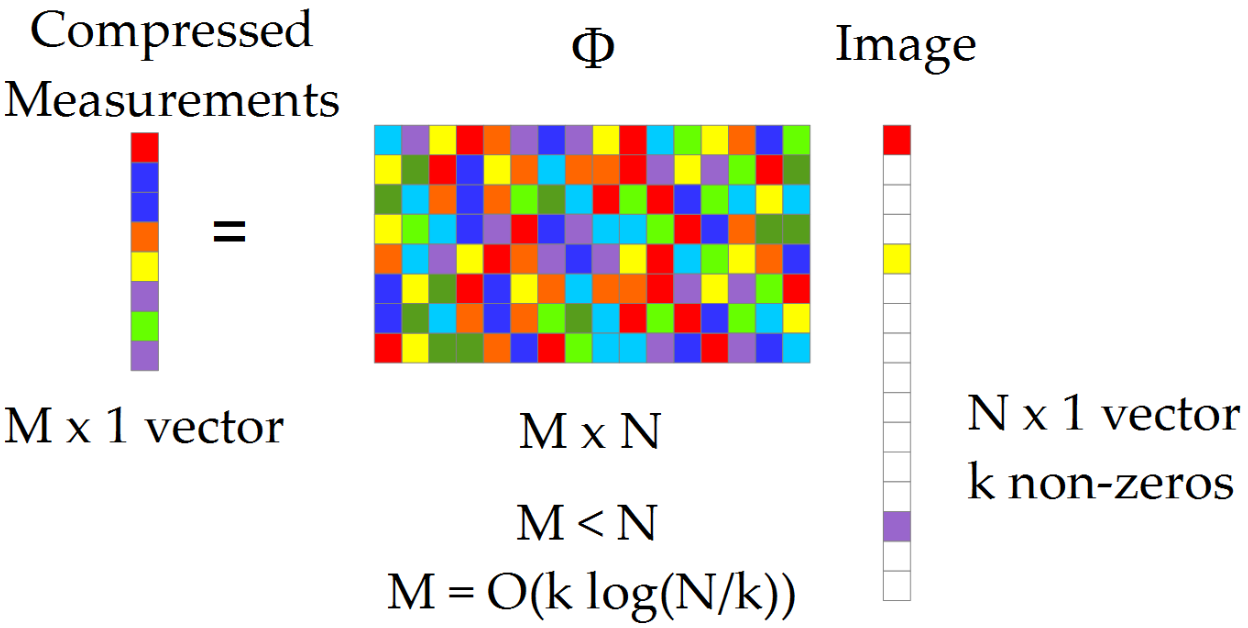
HOG descriptor*



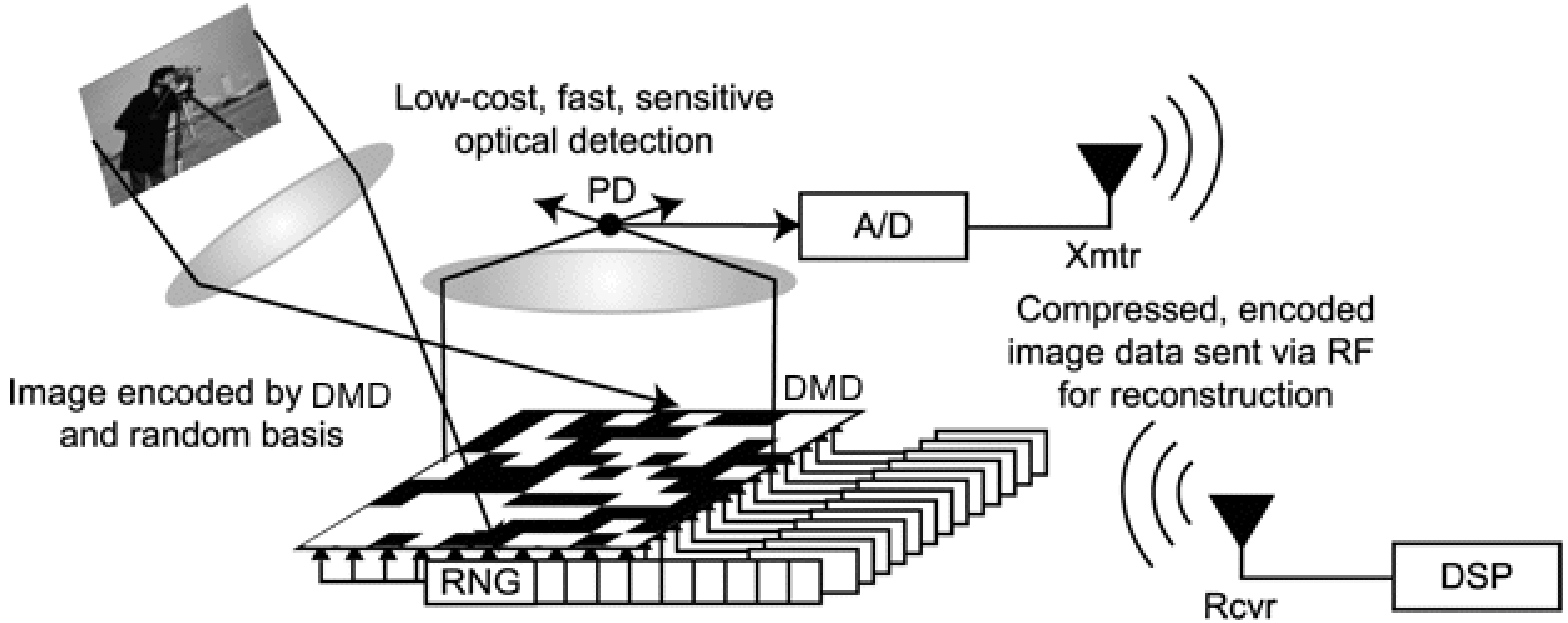
Inference

* N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005

Compressive Imaging



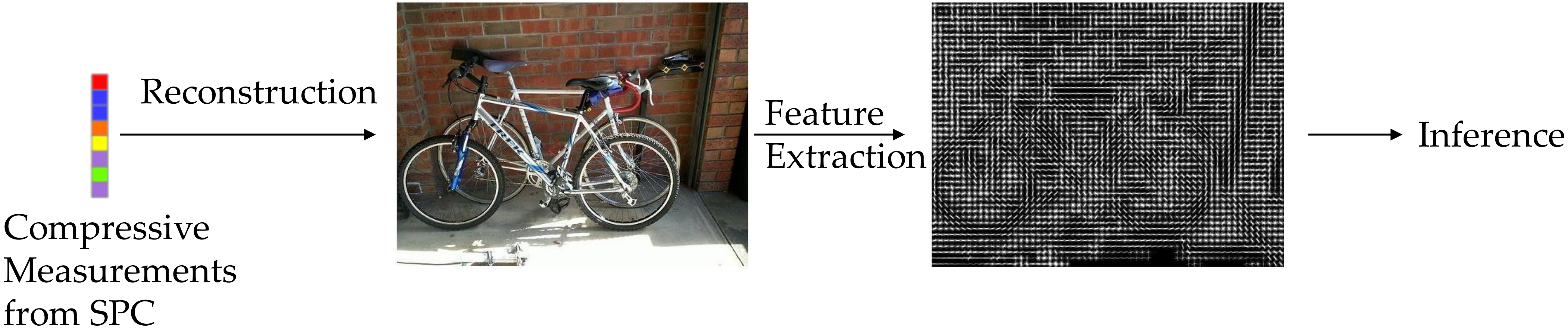
Single Pixel Camera



dsp.rice.edu/cscamera

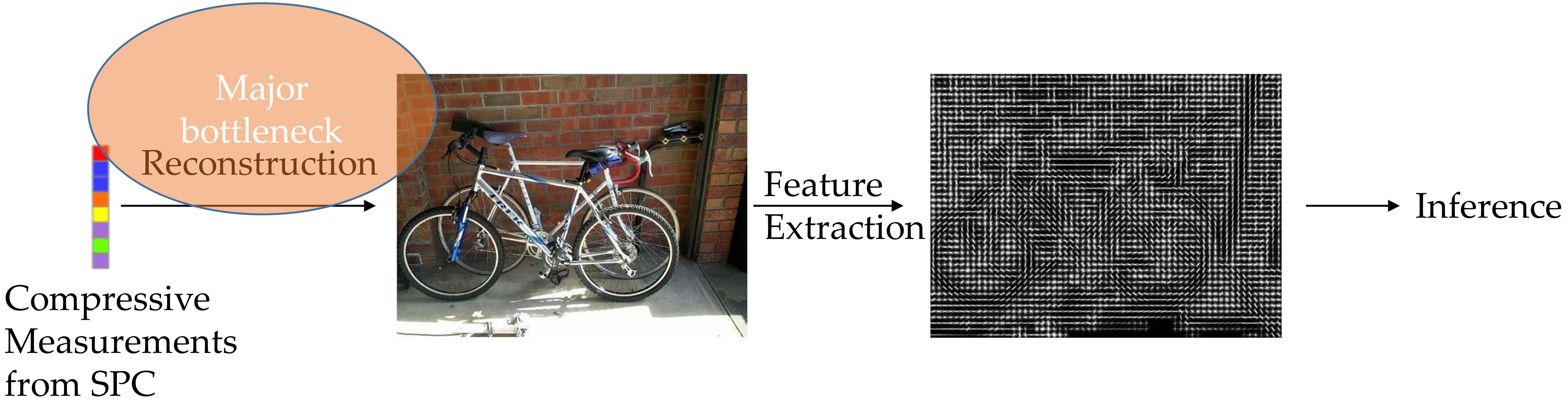
Inference on Compressive Measurements

Reconstruct-then-infer paradigm



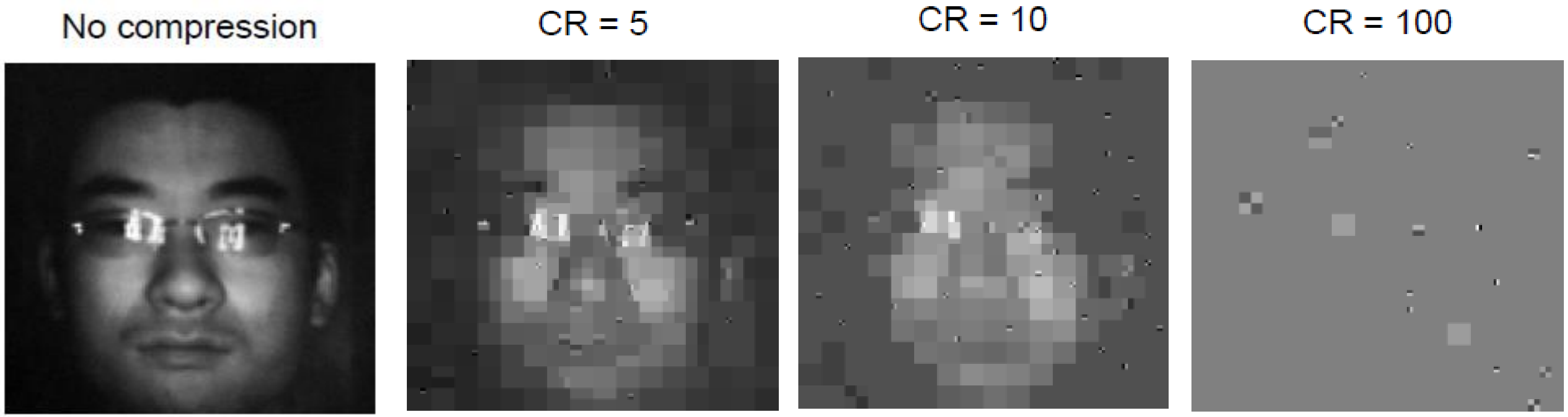
Inference on Compressive Measurements

Reconstruct-then-infer paradigm



Reconstruction algorithms suffer from drawbacks

- Computationally expensive
- Do not produce good results at high compression ratios.

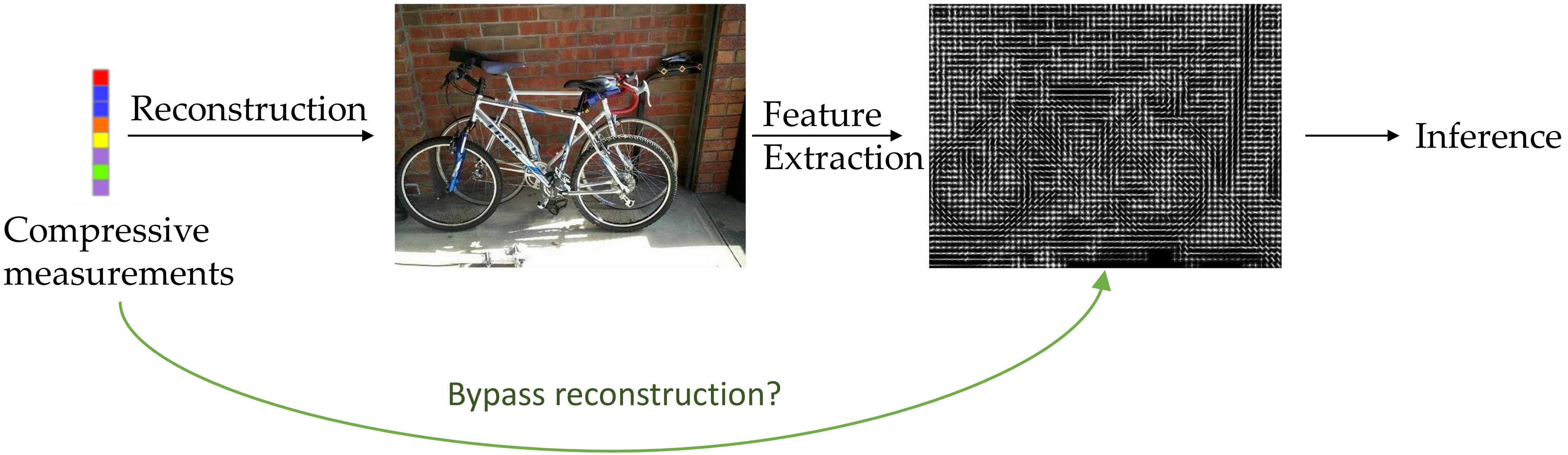


- Various parameters such as sparsity level and sparsifying basis need to be known a priori.

Is Reconstruction Necessary for Inference?

- We posit that one can build effective inference algorithms directly on the compressed bits.

Reconstruction-free compressive inference



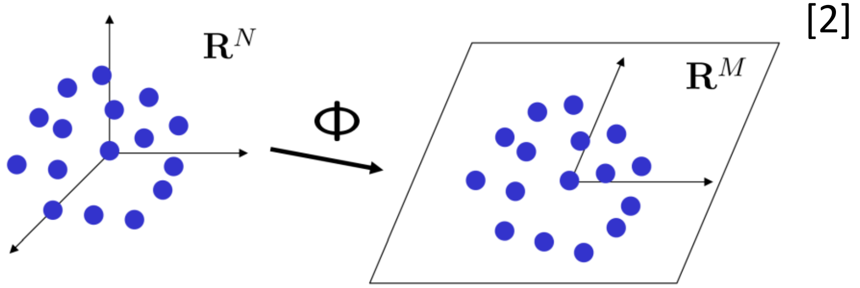
Basis for direct feature extraction

Johnson-Lindenstrauss Lemma [1]

For a given set of Q points in a high dimensional space, certain embeddings exist that nearly preserve distances between points when mapped to a lower dimensional space.

$$f : R^N \to R^M$$

$$M = \mathcal{O}\left(\frac{log(Q)}{\epsilon^2}\right)$$



$$(1-\epsilon)\|v-u\|^2 \leq \|f(v)-f(u)\|^2 \leq (1+\epsilon)\|v-u\|^2$$

- It has been shown that the mapping f can be a random matrix Φ with entries drawn from certain distributions. Such matrices are used in CS.

[1] W.B. Johnson and J. Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space." *Contemporary mathematics* 26.189-206 (1984): 1.

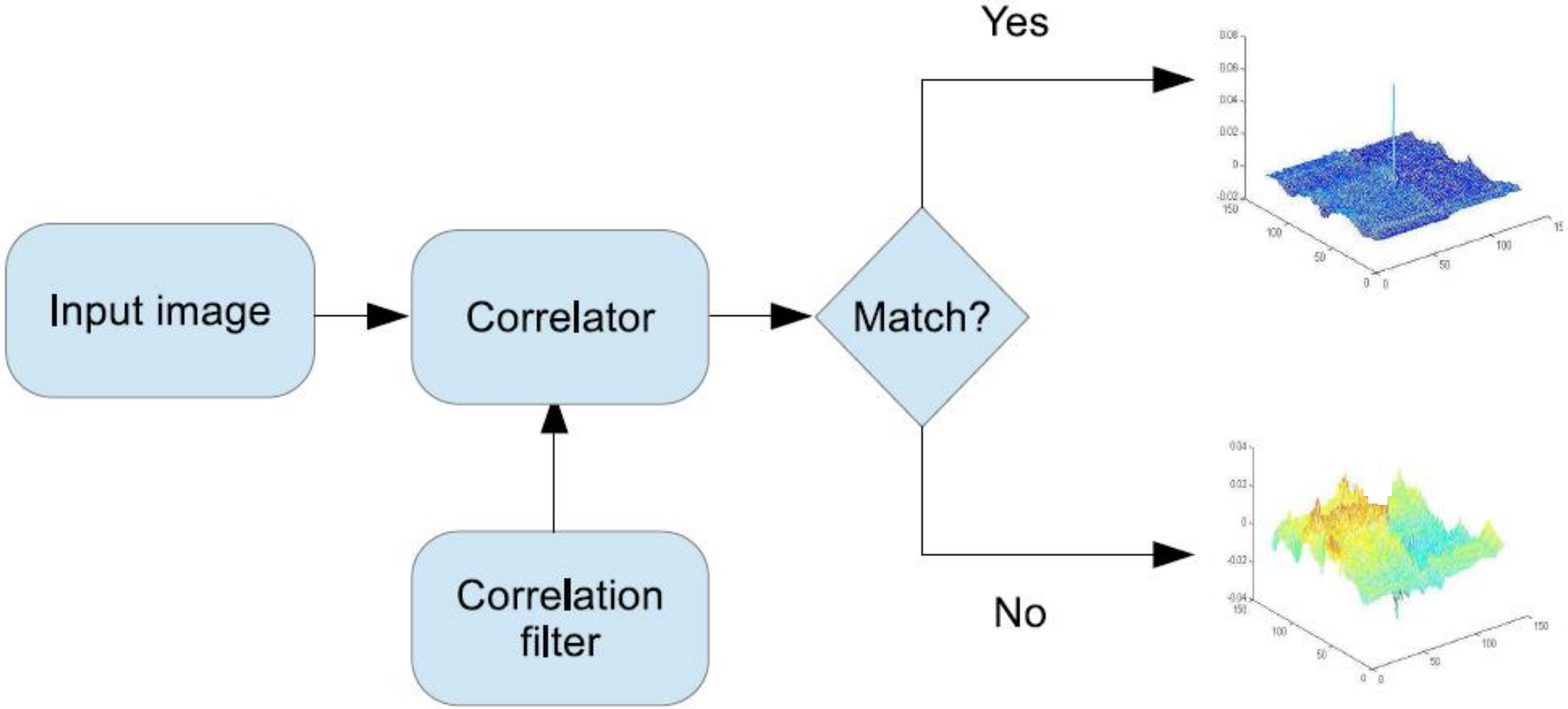
[2] J. Romberg, M. Wakin, "Compressed Sensing: A Tutorial", IEEE Statistical Signal Processing Workshop, 2007.

Basis for direct feature extraction

- This fact is employed in the design of the smashed filter for compressive classification by Davenport et al. (2007) [3].
- We extend this idea to construct smashed correlation filters which are more robust to input variations.
- As a consequence of the JL lemma, inner products and hence, correlations between vectors are also preserved after mapping to a lower dimension.

[3] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly,and R. G. Baraniuk. "The smashed filter for compressive classification and target recognition." In Electronic Imaging, pages 64980H–64980H. International Society for Optics and Photonics, 2007.

Correlation Filters for Visual Recognition



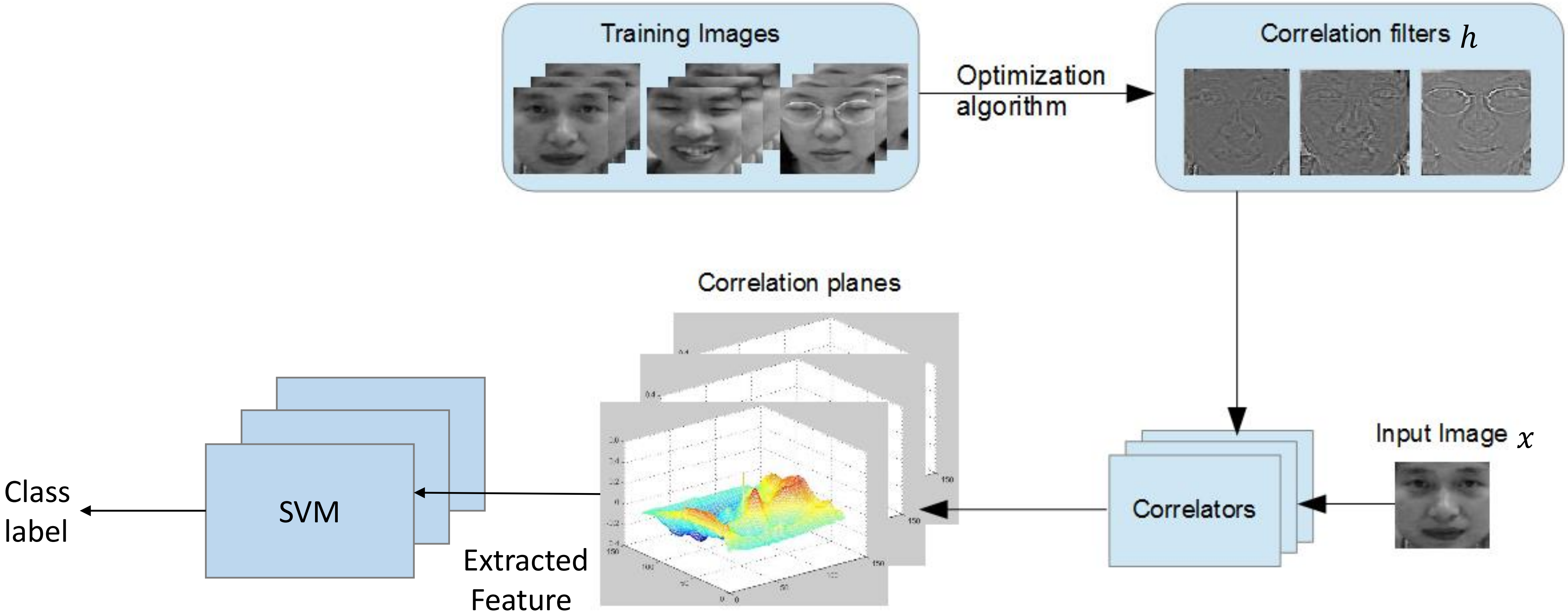
Correlation Filters for Visual Recognition

- For a training set $\{x_1, x_2, \ldots, x_n\}$, g_i being the desired output for image x_i, the correlation filter h* is

$$\mathbf{h}^* = \arg\min_{\mathbf{h}} \quad \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{h} \otimes \mathbf{x}_i - \mathbf{g}_i\|_2^2$$

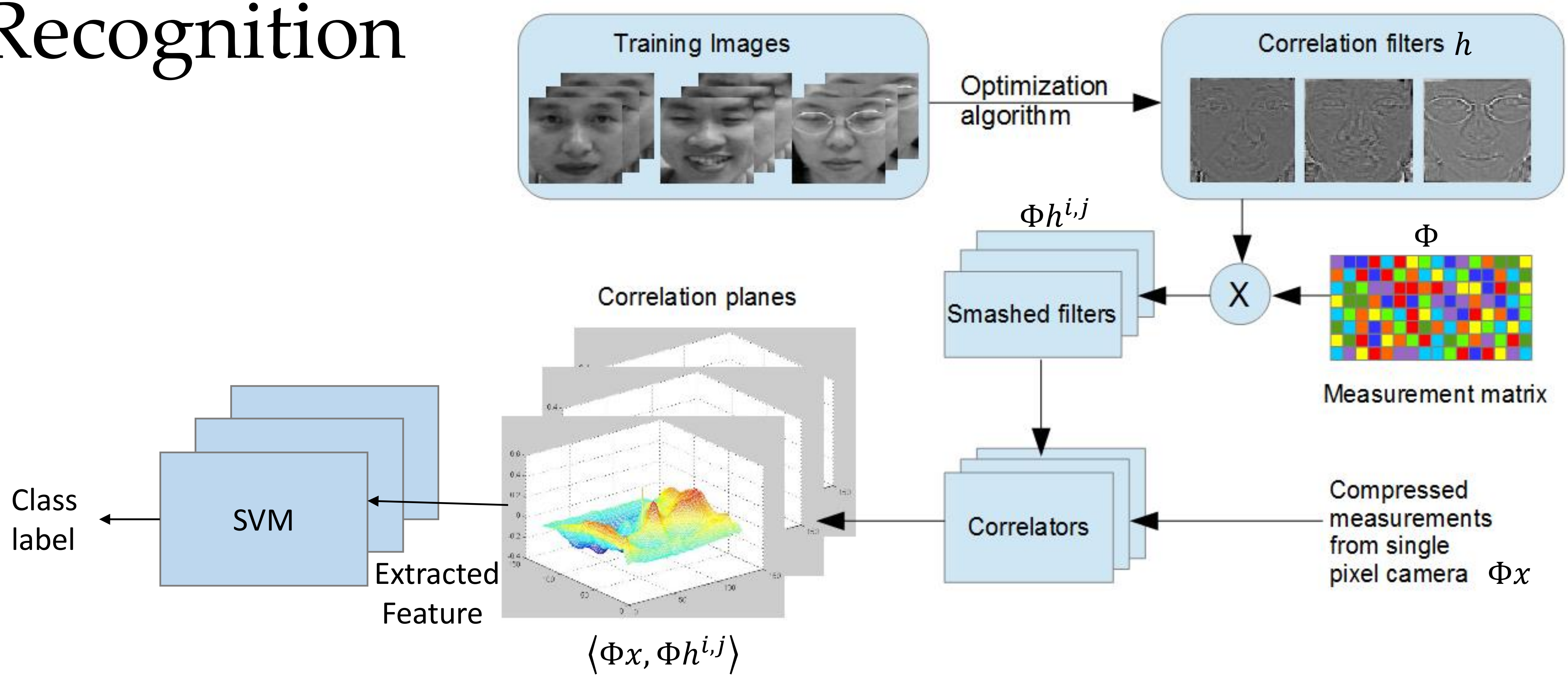
- $g_i = [0\ 0\ \ldots\ 0]$ if x_i belongs to the false class.
- $g_i = [0\ 0\ \ldots 0\ 1\ 0\ \ldots 0]$ if x_i belongs to the true class, the 1 being in the target location.

Recognition with Correlation Filters – Oracle Sensing



$$c(i,j) = \langle x, h^{i,j} \rangle$$

Smashed Correlation Filters for Compressive Recognition



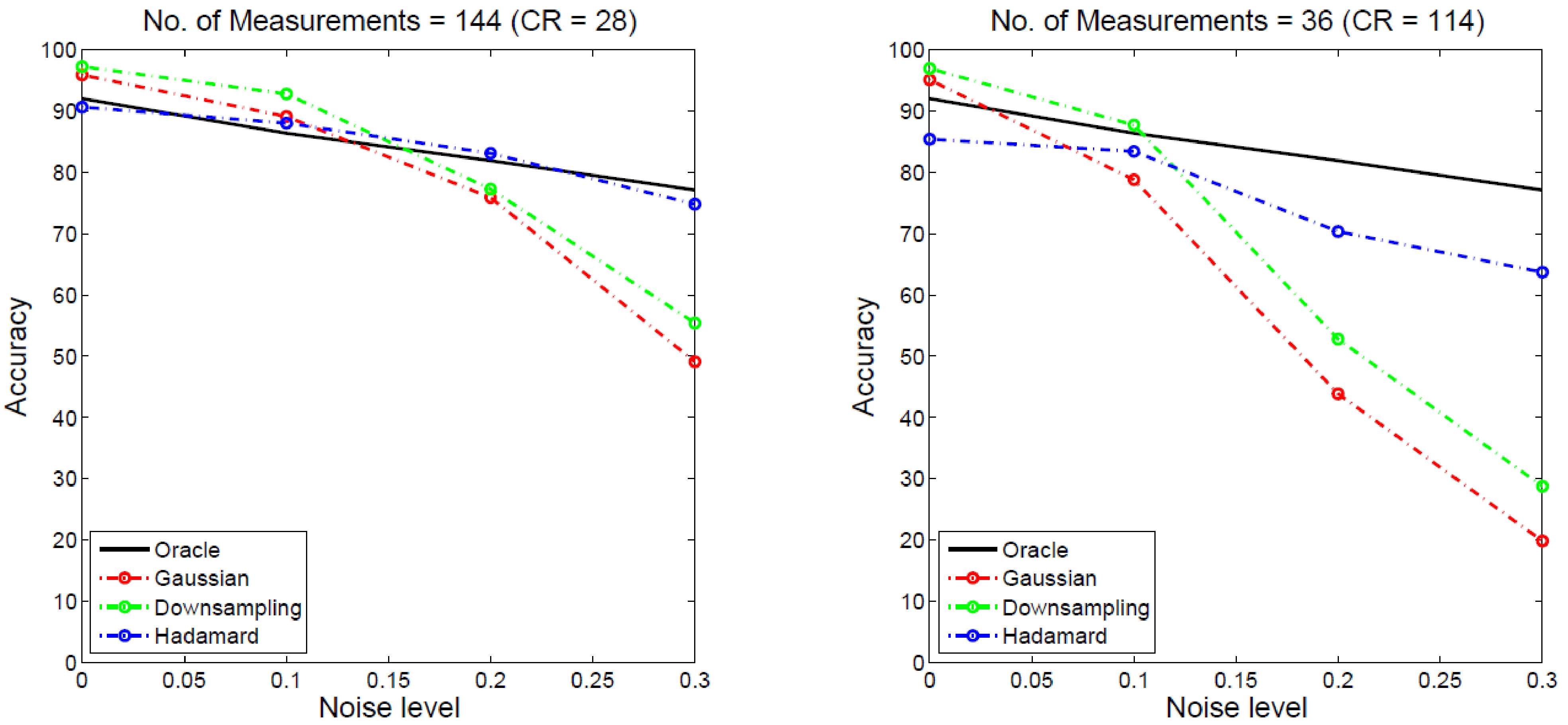
Using a corollary of the JL lemma, we can write $\langle \Phi x, \Phi h^{i,j} \rangle - \epsilon \leq c(i,j) \leq \langle \Phi x, \Phi h^{i,j} \rangle + \epsilon$

Experiments

- Controlled experiments
  1. AMP Database [4]
     - 13 subjects with 75 images per subject.
     - 25 for training, 50 for testing.
     - 64 x 64 images.

[4] http://chenlab.ece.cornell.edu/projects/FaceAuthentication/

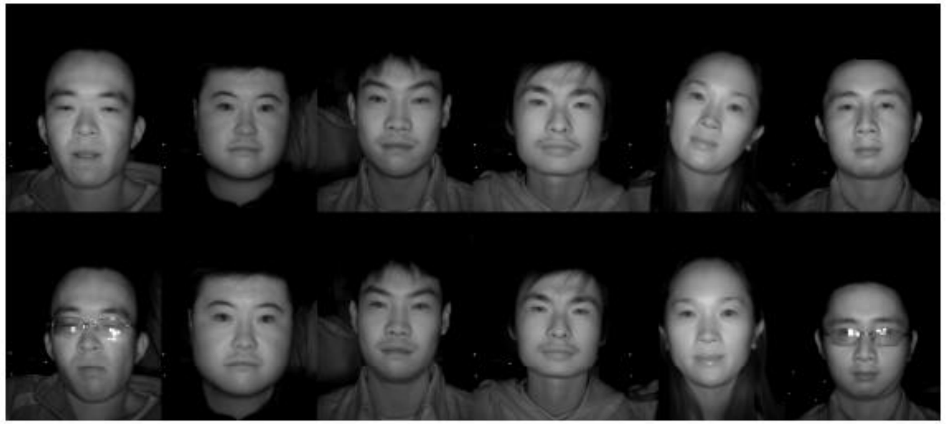
Results on AMP Database (64 x 64 images)



- At low noise + high compression ratios; accuracy is comparable to oracle sensing.
- Hadamard measurements are more robust to noise.

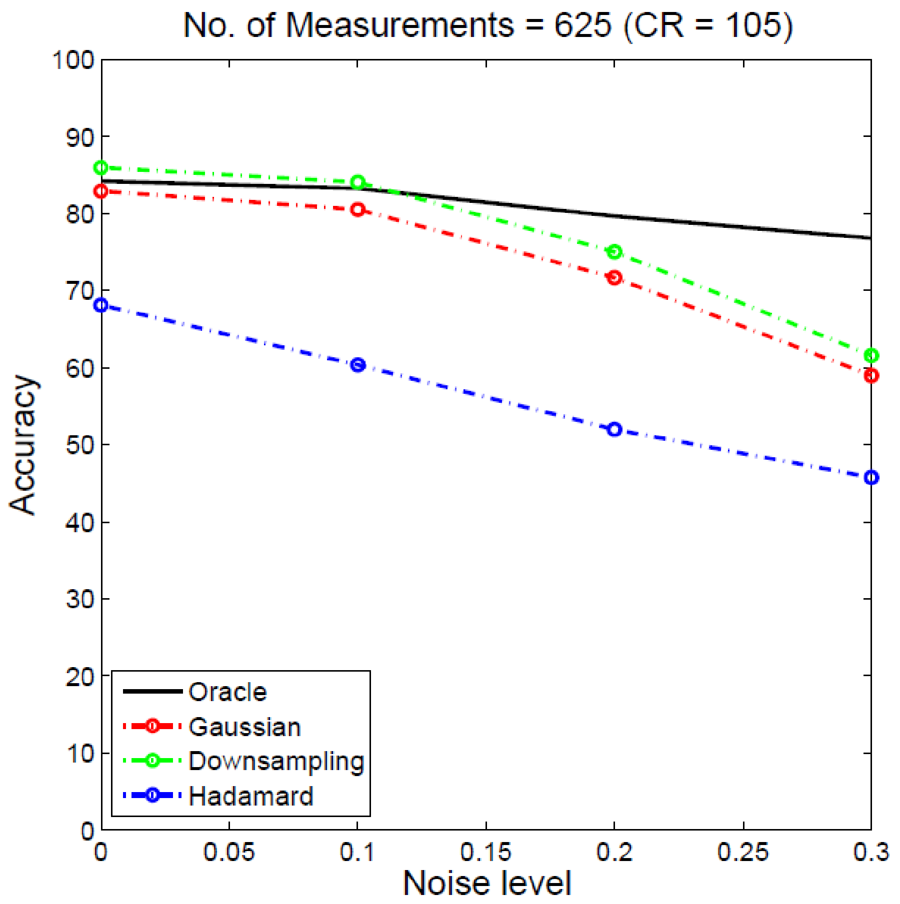
Experiments

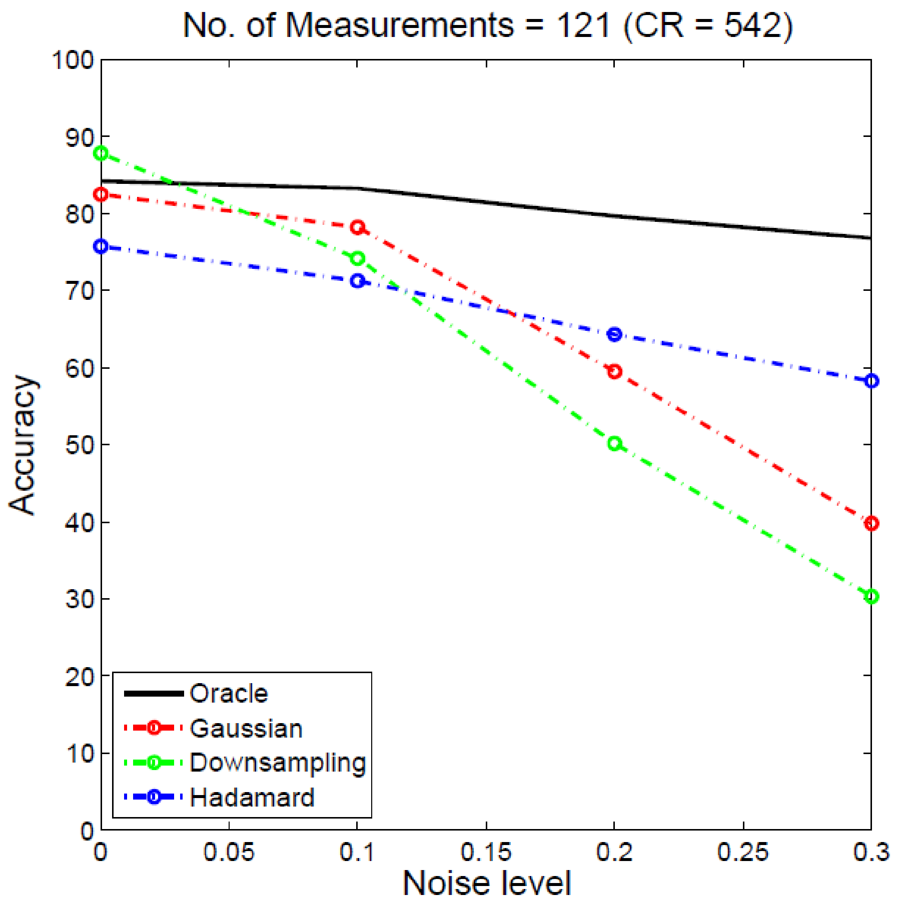
- Controlled experiments
  2. NIR Database [5]
     - Near infrared images.
     - 197 subjects with 20 images per subject.
     - 10 for training, 10 for testing.
     - 640 x 480 images resized to 256 x 256.



[5] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(4):627–639, 2007.

Results on NIR Database (256 x 256 images)

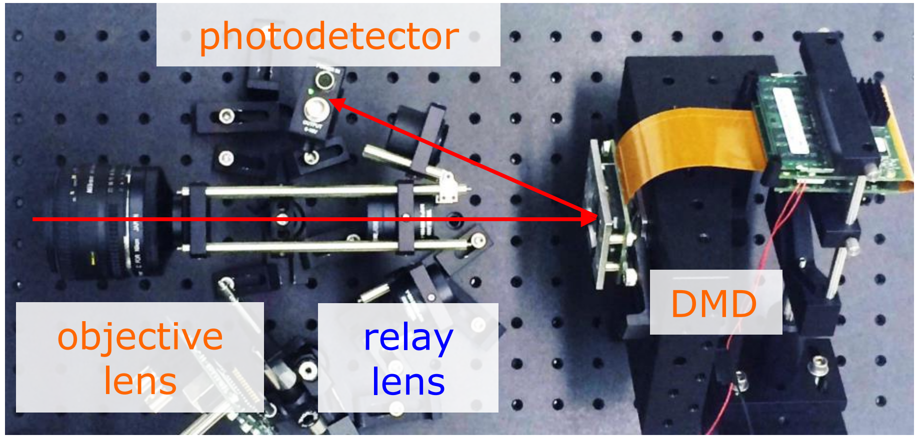




- At low noise + high compression ratios; accuracy is comparable to oracle sensing.
- Hadamard measurements are more robust to noise at high compression ratio.

Experiments on Single Pixel Camera

- New dataset using the SPC consists of CS measurements of 120 facial images of size 128 x 128.
- Images belong to 30 subjects with 4 images per subject.
- Images are captured using a DMD with operating speed of 22,700 measurements per second.
- Hadamard matrix was used for sensing.



Experiments on Single Pixel Camera

- Four training-testing splits were created using the database.
- Each split containing 3 training images and 1 testing image.
- Face recognition accuracy was computed as the average over all the splits.

Face Recognition Results

Compression Ratio	No. of measurements	Recognition Accuracy (%)
1 (Oracle)	16384	60
10	1638	62.5
50	328	58.33
100	164	53.33
200	82	49.17

Benefits of compressive acquisition in IR

- IR imaging provides illumination invariance.
- Non-visible wavelength sensors are expensive.

FLIR T620 IR camera
640 x 480 pixels
Price : $ 21,000

- Compressive imaging -- e.g. SPC -- provides a cost-effective alternative.

Conclusions and Future Work

- We have shown that compressive sensing technology can be employed in a practical inference application by extracting features from compressive measurements directly, using smashed correlation filters, thus avoiding reconstruction.
- Points to new avenues of research for understanding how to solve high-level computer vision problems from computational imagers in general.

References

1. W.B. Johnson, and J. Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space." *Contemporary mathematics* 26.189-206 (1984): 1.
2. J. Romberg, M. Wakin, "Compressed Sensing: A Tutorial", *IEEE Statistical Signal Processing Workshop*, 2007.
3. M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly,and R. G. Baraniuk. “The smashed filter for compressive classification and target recognition.” *Electronic Imaging*, pages 64980H–64980H. International Society for Optics and Photonics, 2007.
4. AMP Database, http://chenlab.ece.cornell.edu/projects/FaceAuthentication
5. S. Z. Li, R. Chu, S. Liao, and L. Zhang. “Illumination invariant face recognition using near-infrared images.” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):627–639, 2007.

Thank you

Supported by ONR and NSF

Preserving Inner Products – Proof Sketch

$f : R^N \to R^M \qquad M = \mathcal{O}(\frac{log(Q)}{\epsilon^2}) \qquad \forall\ u, v \in Q \text{ and } \|u\| = 1, \|v\| = 1,$

Using JL lemma,

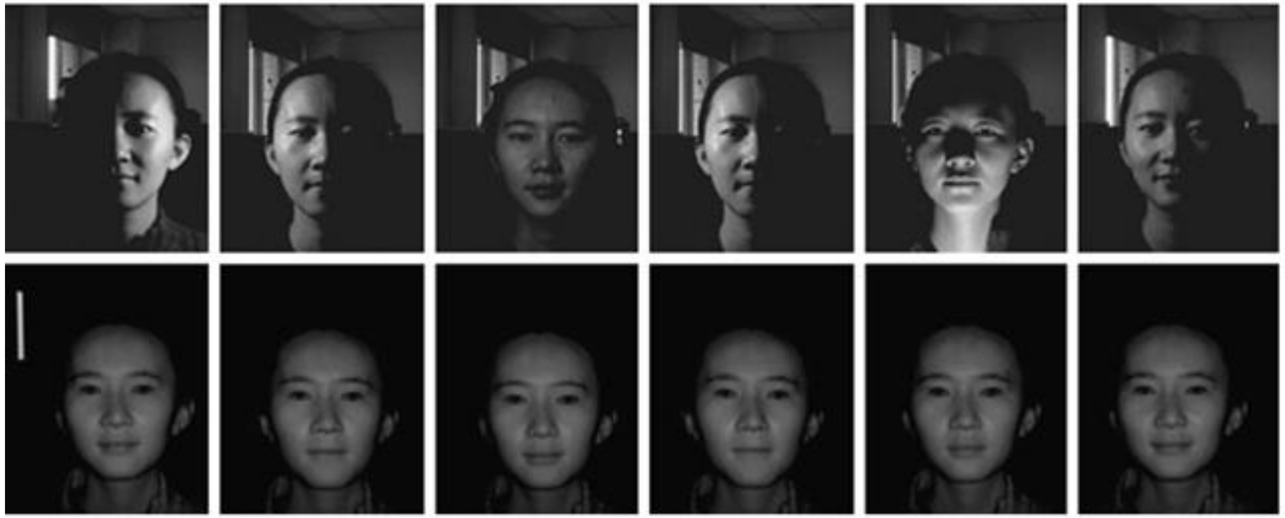
$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2$$
$$(1-\epsilon)\|u+v\|^2 \leq \|f(u)+f(v)\|^2 \leq (1+\epsilon)\|u+v\|^2$$

$$\begin{aligned}
4f(u)\cdot f(v) &= \|f(u)+f(v)\|^2 - \|f(u)-f(v)\|^2 \\
&\geq (1-\epsilon)\|u+v\|^2 - (1+\epsilon)\|u-v\|^2 \\
&= 4u\cdot v - 2\epsilon(\|u\|^2 + \|v\|^2) \\
&= 4u\cdot v - 4\epsilon
\end{aligned}$$

Proving the other direction along the same lines gives us the required result.

Compressive Sensing for Inference

Benefits of IR imaging for inference – illumination invariance



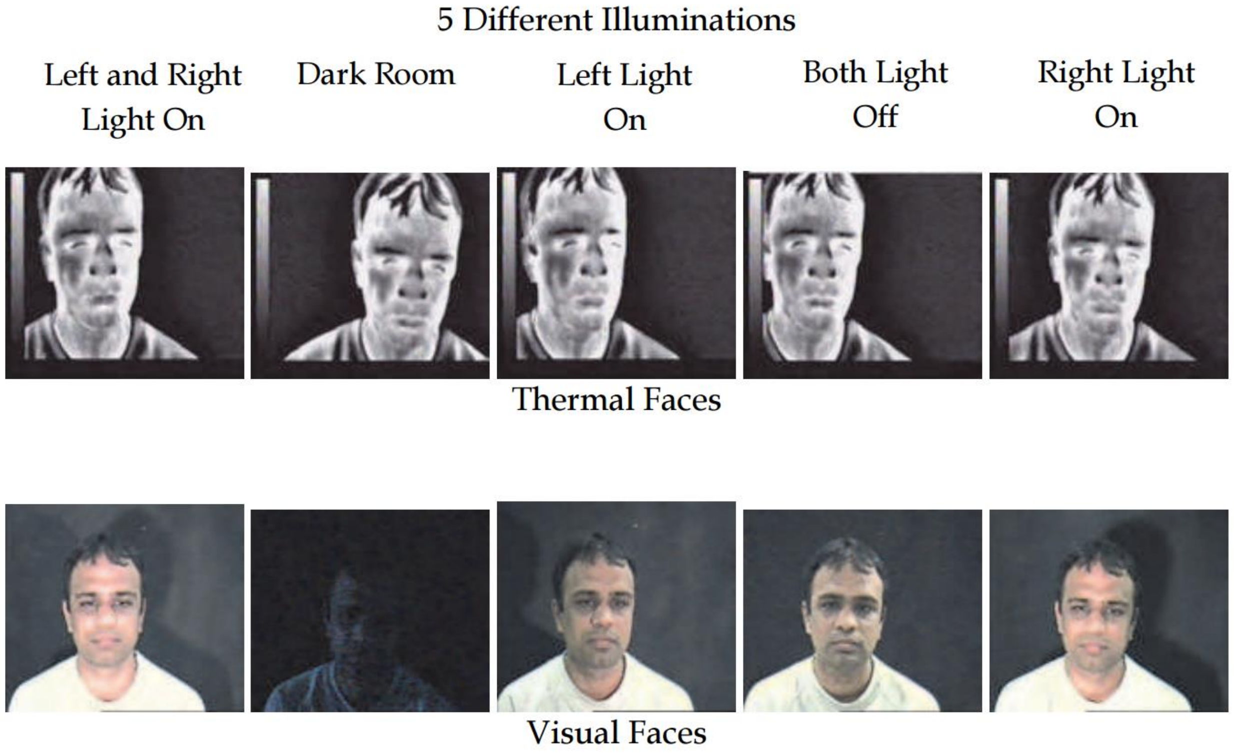
Color camera

Near-infrared imager

S. Z. Li, R. Chu, S. Liao, and L. Zhang. "Illumination invariant face recognition using near-infrared images." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):627–639, 2007.

Compressive Sensing for Inference

Benefits of IR imaging for inference – illumination invariance



Long-wave IR imager

Color camera

Acquisition Time Can Be Significantly Reduced

- Our SPC senses 22.7 measurements/second
- For a 128 x 128 image, sensing time is

Sensing Mechanism	% Measurements Required	Sensing Time
Sensing all measurements	100% (16384)	0.72 s
Compressive sensing with recovery	20%	~ 0.1 s
Compressive sensing without reconstruction	1%	~ 0.001 s

Maximum Margin Correlation Filters

- Developed by Rodriguez et al. (2013)
- Combines the strengths of correlation filters and SVM.
- Optimize for
  - High peak at target location.
  - Maximum margin (SVM objective)
- Solve optimization problem of the form:

$$\min_{\mathbf{h},b} \quad \left(\|\mathbf{h}\|_2^2 + C\sum_{i=1}^{N}\xi_i, \sum_{i=1}^{N}\|\mathbf{h}\otimes\mathbf{x}_i - \mathbf{g}_i\|_2^2\right)$$

$$\text{s.t.} \quad t_i(\mathbf{h}^T\mathbf{x_i} + b) \geq c_i - \xi_i, \quad i = 1, 2, \ldots, N$$

- Can be reduced to a single optimization problem that can be solved on any standard SVM solver.

A. Rodriguez, V. N. Boddeti, B. V. Kumar, and A. Mahalanobis. Maximum margin correlation filter: A new approach for localization and classification. IEEE Transactions on Image Processing, 22(2):631–643, 2013.

Feature extraction from correlation planes

- Each correlation plane is divided into non-overlapping blocks and the PSR and peak values of each block is extracted.
- The peak and PSR for the entire correlation plane are also extracted.
- All these values are concatenated to form the feature vector that is input to the SVMs.

$$\hat{c}(i,j) = \langle \Phi^T \Phi \mathbf{x}, H^{i,j} \rangle.$$