3D²PM - 3D Deformable Part Models

Bojan Pepik[1] Peter Gehler[2] Michael Stark[1,3] Bernt Schiele[1]

[1]Max Planck Institute for Informatics, Saarbrücken, Germany [2]Max Planck Institute for Intelligent Systems, Tübingen, Germany [3]Stanford University

Motivation

- Objects are inherently ***3-dimensional***
- 3D object representations provide:
  - ***Compact*** and ***accurate*** approximation of the physical world
- Higher level vision tasks can benefit from ***expressive*** object detectors:
  - ***Angular accurate*** viewpoints
  - ***3D parts*** consistent across views
- *State-of-the-art* detectors are modeled in 2D
- 3D object detectors lack detection performance

Contributions

➡ ***3D version*** of the Deformable Part Model [2] capable of:
- Richer object hypotheses (beyond 2D BB)
- Robust matching to image evidence

➡ ***Richer*** object hypotheses:
- Viewpoint estimation of arbitrary granularity
- Consistent parts across views

➡ ***Favorable*** performance:
- *State-of-the-art* viewpoint estimation results
- Competitive 2D object localization results

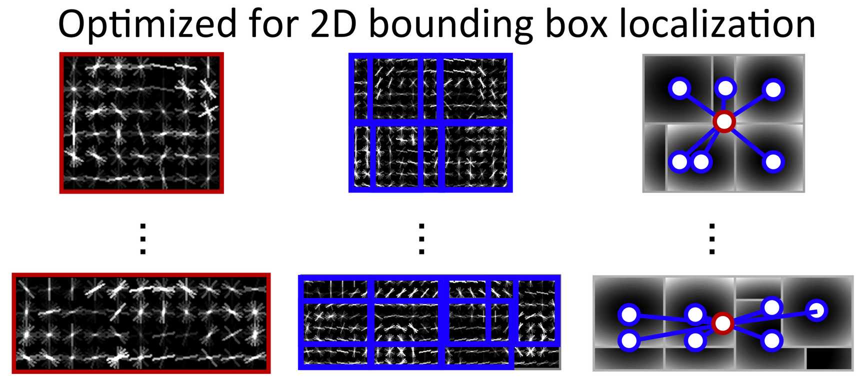
➡ ***Jointly optimize*** for object localization and continuous viewpoint estimation

3D pose estimation results

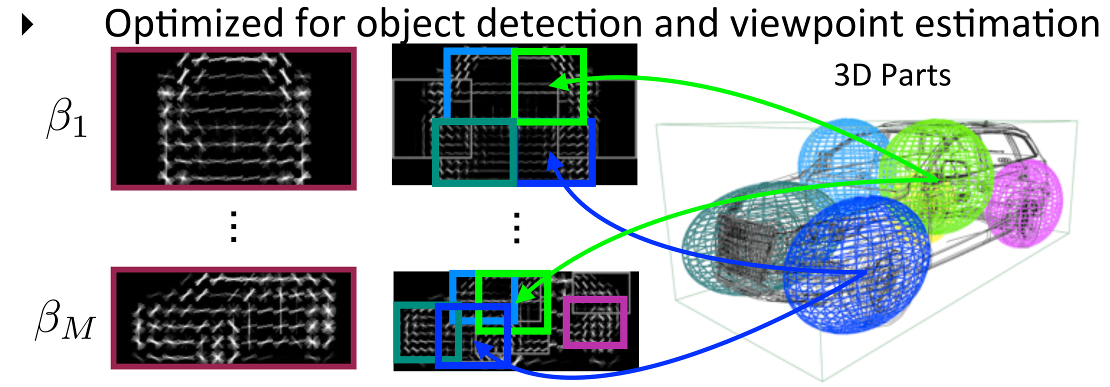
***Note the color coded part correspondences**

3D Deformable Part Models

- Deformable part model [2]
  - Mixture of star CRFs in 2D space
  - Parts are independent across components
  - Discrete appearance model
  - Optimized for 2D bounding box localization

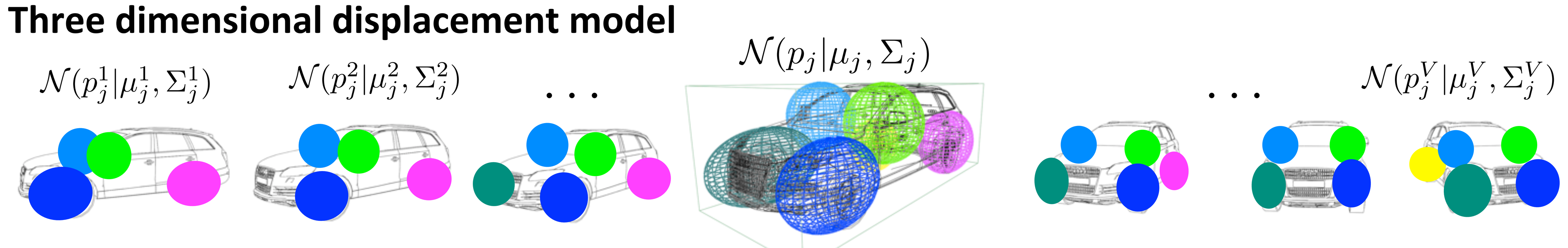


- 3D²PM
  - A star CRF in 3D space
  - Parts are in 3D $p_j = (x_j, y_j, z_j)$ and linked
  - Continuous appearance model
  - Optimized for object detection and viewpoint estimation



3D Parts

Three dimensional displacement model



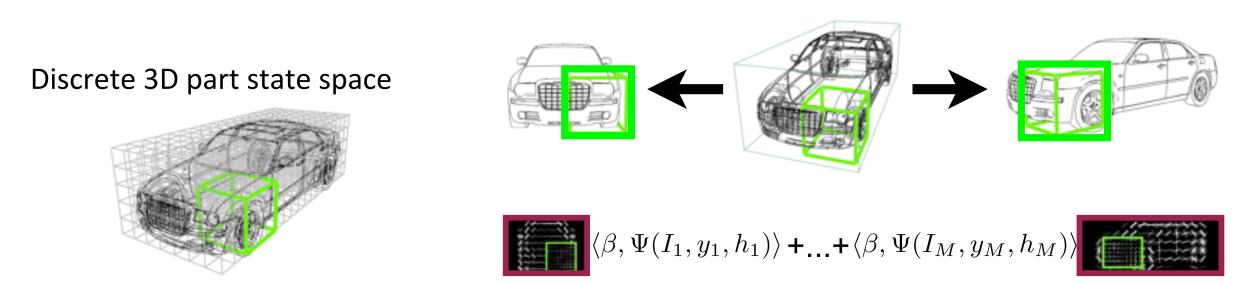
- **Compact** part displacement distribution in 3D $p(p_j|p_o) = \mathcal{N}(p_j|\mu_j, \Sigma_j)$
  - ***6*** parameters per part compared to ***4M*** in DPM [1,2]
- In view projected part distribution

orthographic projection

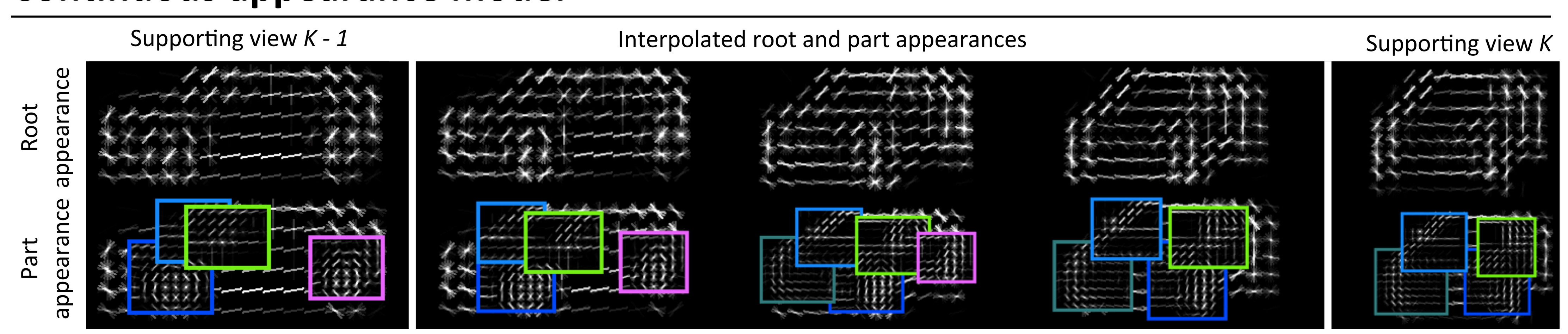
$$p(p_j^v|p_o^v) = \mathcal{N}(p_j^v|Q_j^v\mu_j, (Q_j^v)\Sigma_j(Q_j^v)^T)$$

3D part inference

- Part inference in 3D per object instance
  - Across all views of a given object instance in which the part is visible in
  - Projected parts are observed by viewpoint-specific model instantiations



Continuous appearance model



- Linear and exponential appearance interpolation scheme
- The model can synthesize ***infinitely many components*** without the need to learn them all
- Allows arbitrary fine viewpoint estimation
- Faster inference w.r.t. to brute-force

Model training

- Structured output SVM with margin rescaling
- Jointly address object localization and viewpoint estimation

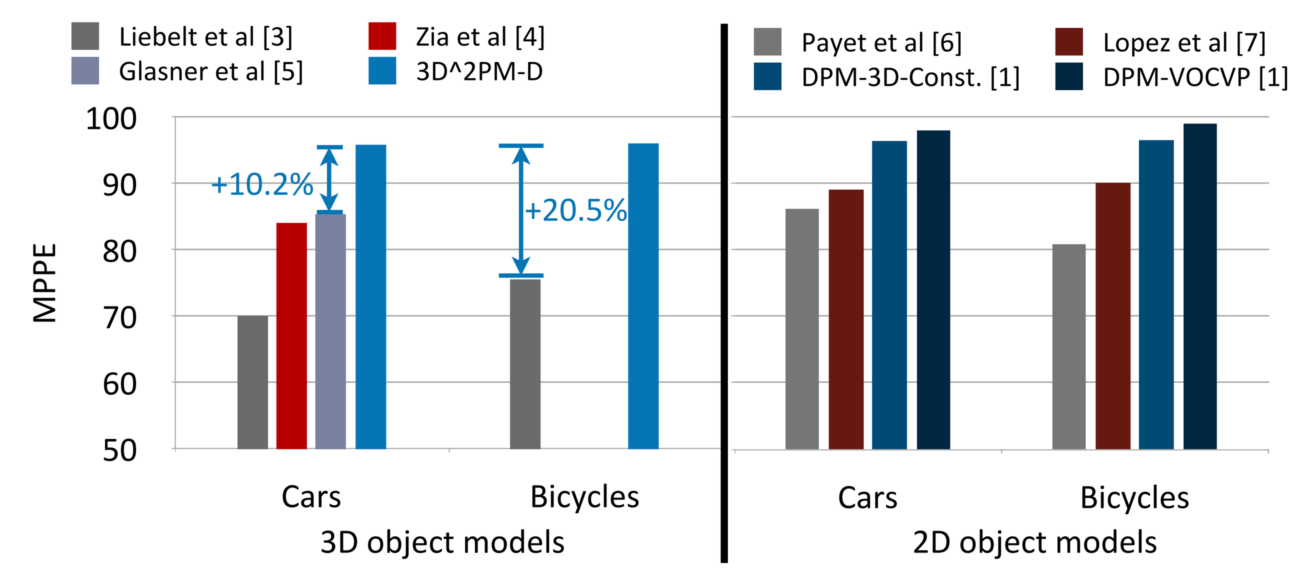
$$\Delta(y, \bar{y}) = \alpha\Delta_{VOC}(y, \bar{y}) + (1-\alpha)\Delta_{VP}(y, \bar{y})$$

$$\Delta_{VOC}(y, \bar{y}) = 1 - \frac{y^b \cap \bar{y}^b}{y^b \cup \bar{y}^b} \quad \Delta_{VP}(y, \bar{y}) = \frac{\angle(y^v, \bar{y}^v)}{180^\circ}$$

Experiments

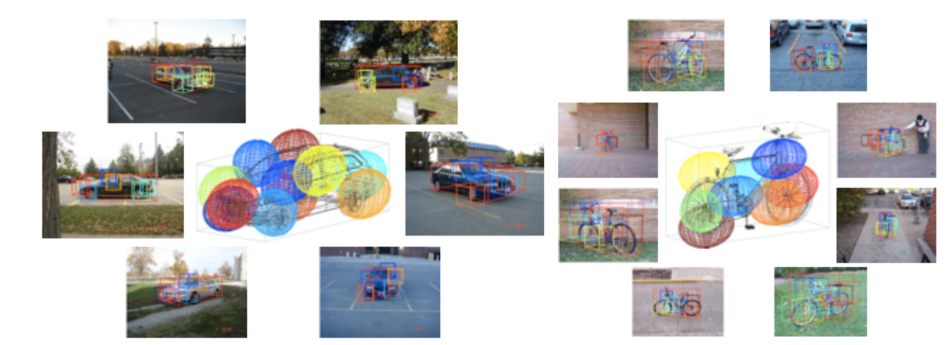
Coarse-grained viewpoint estimation (viewpoint classification)

3D OBJECT CLASSES DATASET (8 discrete viewpoint classes)



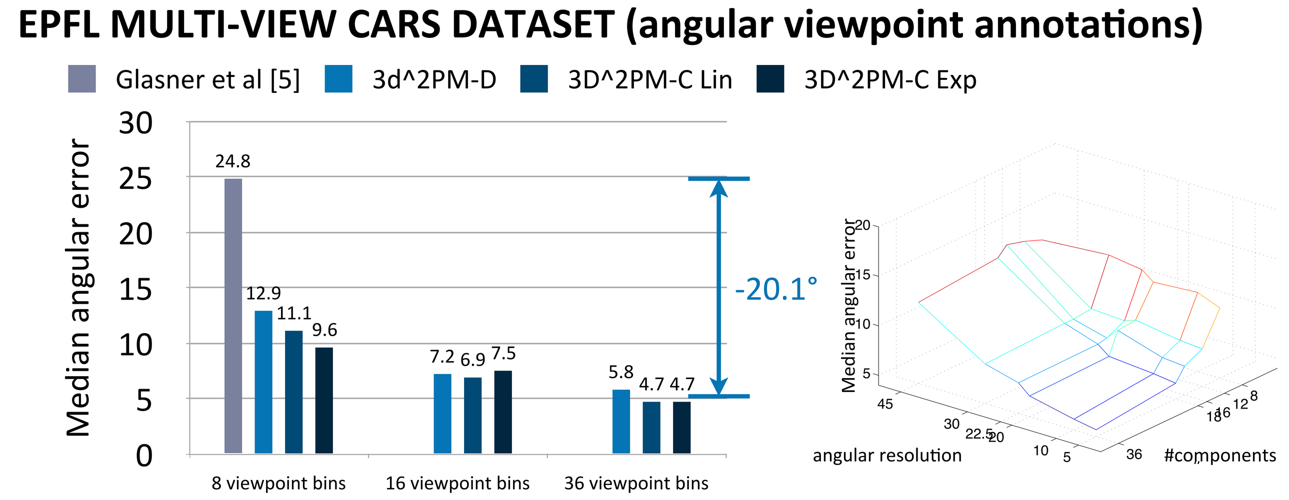
➡ 3D²PM model ***outperforms*** 3D object models by large margin

➡ ***On par*** performance to 2D object models



Fine-grained viewpoint estimation (angular viewpoint estimation)

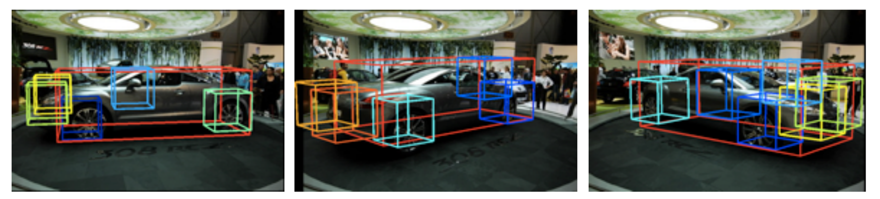
EPFL MULTI-VIEW CARS DATASET (angular viewpoint annotations)



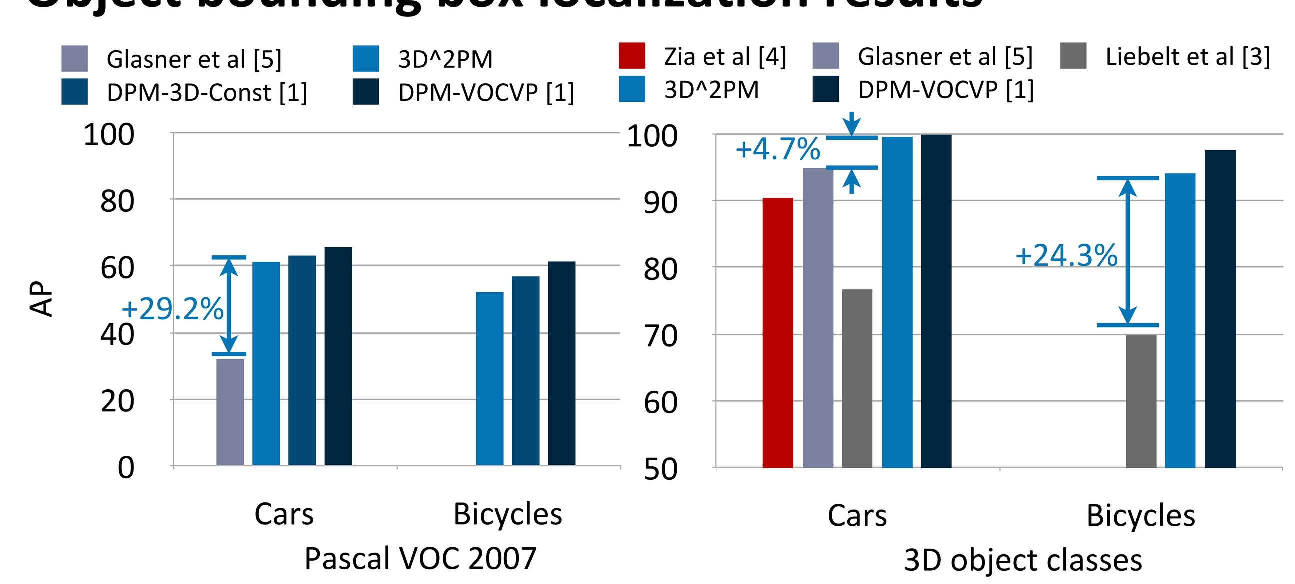
➡ State-of-the-art multi-view models can output discrete set of viewpoints

➡ Our ***3D²PM-C*** can provide fine viewpoint estimates

➡ 3D²PM-C ***state-of-the-art*** on angular viewpoint estimation



Object bounding box localization results



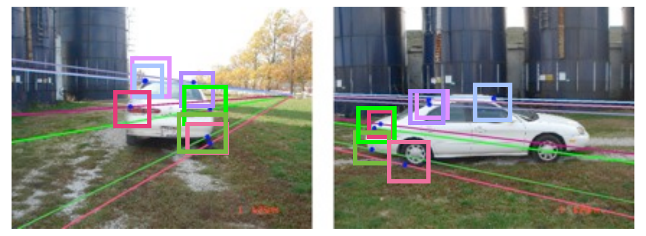
➡ Unlike the state-of-the-art 3D models, 3D²PMs achieve detection performance on par with the 2D object models

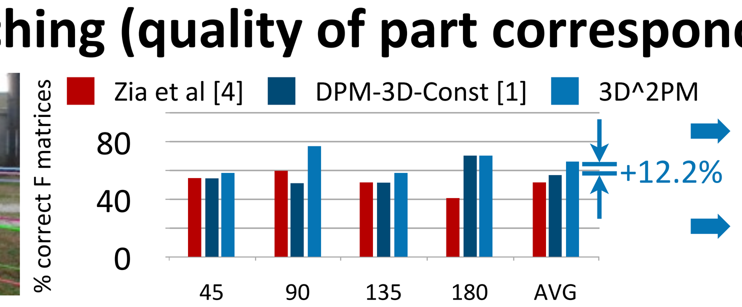
- **Coarse to fine inference**

AP /MAE	at 5°	#atomic operations
3D²PM-C b36 full inference	99.2 / 4.7	2.20×10^{10}
3D²PM-C b36 coarse to fine	**99.0 / 7.0**	**0.48×10^{10}**
3D²PM-C b12	97.6 / 7.5	2.20×10^{10}
3D²PM-C b18	98.0 / 6.9	2.20×10^{10}

➡ x5 faster inference at minimal loss

Ultra-wide baseline matching (quality of part correspondences)





➡ 3D²PM due to the 3D model gives accurate part correspondences

➡ 3D²PM ***state-of-the-art*** on ultra-wide baseline matching experiment

References

[1] B. Pepik, M. Stark, P. Gehler, B. Schiele Teaching 3D Geometry to Deformable Part models CVPR'12
[2] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan Object Detection with Discriminatively Trained Part Based Models PAMI'10
[3] J. Liebelt, C. Schmid Multi-view Object Class Detection With A 3D Geometric Model CVPR'10
[4] M. Z. Zia, M. Stark, B. Schiele, and K. Schindler. Revisiting 3D Geometric Models for Accurate and Object Shape and Pose 3DRR'11
[5] D. Glasner, M. Galun, S. Alpert, R. Basri, G. Shakhnarovich Viewpoint-Aware Object Detection and Pose Estimation ICCV'11
[6] N. Payet, S. Todorovic, From Contours to 3D Object Detection and Pose Estimation ICCV'11
[7] R. J. López-Sastre, T. Tuytelaars, S. Savarese. Deformable Part Models Revisited: A Performance Evaluation for Object Category Pose Estimation CORP'11

Acknowledgements: This work has been supported by the Max Planck Center for Visual Computing and Communication. We thank M. Zeeshan Zia for his help in conducting wide baseline matching experiments.