<table>
<thead>
<tr>
<th>Oral 1A / Poster 1</th>
<th>Topic: 3D Vision and Computational Photography</th>
</tr>
</thead>
<tbody>
<tr>
<td>132</td>
<td>Inferring Super-Resolution Depth from a Moving Light-Source Enhanced RGB-D Sensor: A Variational Approach</td>
</tr>
<tr>
<td>250</td>
<td>Unsupervised Learning of Camera Pose with Compositional Re-estimation</td>
</tr>
<tr>
<td>413</td>
<td>Blended Convolution and Synthesis for Efficient Discrimination of 3D Shapes</td>
</tr>
<tr>
<td>446</td>
<td>A Multi-Scale Guided Cascade Hourglass Network for Depth Completion</td>
</tr>
<tr>
<td>502</td>
<td>Silhouette Guided Point Cloud Reconstruction beyond Occlusion</td>
</tr>
<tr>
<td>516</td>
<td>Non-Rigid Structure from Motion: Prior-Free Factorization Method Revisited</td>
</tr>
<tr>
<td>616</td>
<td>PointGrowNet: An Autoregressive Approach towards Point Cloud Generation</td>
</tr>
<tr>
<td>691</td>
<td>Depth Completion via Deep Basis Fitting</td>
</tr>
<tr>
<td>697</td>
<td>High Accuracy Face Geometry Capture using a Smartphone Video</td>
</tr>
<tr>
<td>922</td>
<td>FlowNet3D++: Geometric Losses For Deep Scene Flow Estimation</td>
</tr>
<tr>
<td>35</td>
<td>Style Transfer for Light Field Photography</td>
</tr>
<tr>
<td>130</td>
<td>Fourier Based Pre-Processing For Seeing Through Water</td>
</tr>
<tr>
<td>566</td>
<td>DeOccNet: Learning to See Through Foreground Occlusions in Light Fields</td>
</tr>
<tr>
<td>567</td>
<td>Shape from Water Reflection</td>
</tr>
<tr>
<td>686</td>
<td>An Extended Exposure Fusion and its Application to Single Image Contrast Enhancement</td>
</tr>
<tr>
<td>972</td>
<td>Online Lens Motion Smoothing for Video Autofocus</td>
</tr>
<tr>
<td>1086</td>
<td>Fast Image Reconstruction with an Event Camera</td>
</tr>
<tr>
<td>339</td>
<td>Self-Guided Novel View Synthesis via Elastic Displacement Network</td>
</tr>
<tr>
<td>369</td>
<td>On Scene Flow Computation of Gas Structures with Optical Gas Imaging Cameras</td>
</tr>
<tr>
<td>716</td>
<td>Fast Deep Stereo with 2D Convolutional Processing of Cost Signatures</td>
</tr>
<tr>
<td>875</td>
<td>Triple-SGM: Stereo Processing using Semi-Global Matching with Cost Fusion</td>
</tr>
<tr>
<td>67</td>
<td>Optimizing Through Learned Errors for Accurate Sports Field Registration</td>
</tr>
<tr>
<td>376</td>
<td>Reference Grid-assisted Network for 3D Point Signature Learning from Point Clouds</td>
</tr>
<tr>
<td>79</td>
<td>Stable Intrinsic Auto-Calibration from Fundamental Matrices of Devices with Uncorrelated Camera Parameters</td>
</tr>
<tr>
<td>924</td>
<td>Deep Image Blending</td>
</tr>
<tr>
<td>1241</td>
<td>Hand-Priming in Object Localization for Assistive Egocentric Vision</td>
</tr>
<tr>
<td>104</td>
<td>Exploiting Geometric Constraints on Dense Trajectories for Motion Saliency</td>
</tr>
<tr>
<td>Oral 1B / Poster 1</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Topic: Biometrics</strong></td>
<td></td>
</tr>
<tr>
<td>291</td>
<td>Towards a General Model of Knowledge for Facial Analysis by Multi-Source Transfer Learning</td>
</tr>
<tr>
<td>347</td>
<td>Cross-Domain Face Synthesis using a Controllable GAN</td>
</tr>
<tr>
<td>393</td>
<td>Does Face Recognition Accuracy Get Better With Age? Deep Face Matchers Say No</td>
</tr>
<tr>
<td>473</td>
<td>Appearance-Based Gaze Estimation via Gaze Decomposition and Single Gaze Point Calibration</td>
</tr>
<tr>
<td>612</td>
<td>Detecting Morphed Face Attacks Using Residual Noise from Deep Multi-Scale Context Aggregation Network</td>
</tr>
<tr>
<td>680</td>
<td>Gaze Estimation for Assisted Living Environments</td>
</tr>
<tr>
<td>750</td>
<td>On Hallucinating Context and Background Pixels from a Face Mask using Multi-scale GANs</td>
</tr>
<tr>
<td>759</td>
<td>EyeGAN: Gaze-Preserving, Mask-Mediated Eye Image Synthesis</td>
</tr>
<tr>
<td>870</td>
<td>Boosting Deep Face Recognition via Disentangling Appearance and Geometry</td>
</tr>
<tr>
<td>1094</td>
<td>Robust Facial Landmark Detection via Aggregation on Geometrically Manipulated Faces</td>
</tr>
<tr>
<td>1141</td>
<td>End to End Lip Synchronization with a Temporal AutoEncoder</td>
</tr>
<tr>
<td>1199</td>
<td>Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Recognition?</td>
</tr>
<tr>
<td>1246</td>
<td>AutoToon: Automatic Geometric Warping for Face Cartoon Generation</td>
</tr>
<tr>
<td>1272</td>
<td>Component Attention Guided Face Super-Resolution Network: CAGFace</td>
</tr>
<tr>
<td>962</td>
<td>Nonparametric Structure Regularization Machine for 2D Hand Pose Estimation</td>
</tr>
<tr>
<td>61</td>
<td>3D Hand Pose Estimation with Disentangled Cross-Modal Latent Space</td>
</tr>
<tr>
<td>89</td>
<td>Robust Template-Based Non-Rigid Motion Tracking Using Local Coordinate Regularization</td>
</tr>
<tr>
<td>93</td>
<td>DGGAN: Depth-image Guided Generative Adversarial Networks for Disentangling RGB and Depth Images for 3D Hand Pose Estimation</td>
</tr>
<tr>
<td>332</td>
<td>Multiview Supervision By Registration</td>
</tr>
<tr>
<td>346</td>
<td>DeepFuse: An IMU-Aware Network for Real-Time 3D Human Pose Estimation from Multi-View Image</td>
</tr>
<tr>
<td>626</td>
<td>Attention-based Fusion for Multi-source Human Image Generation</td>
</tr>
<tr>
<td>660</td>
<td>A Tracking Module for Pose Estimation using Data Assimilation Methods</td>
</tr>
<tr>
<td>880</td>
<td>Reducing Footskate in Human Motion Reconstruction with Ground Contact Constraints</td>
</tr>
<tr>
<td>903</td>
<td>Unsupervised Cross-Dataset Adaptation Via Probabilistic Amodal 3D Human Pose Completion</td>
</tr>
<tr>
<td>1004</td>
<td>Lightweight 3D Human Pose Estimation Network Training Using Teacher-Student Learning</td>
</tr>
<tr>
<td>Awards Session / Poster 2</td>
<td>Oral 2A / Poster 2</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Topic: Action Recognition</strong></td>
<td></td>
</tr>
<tr>
<td>TBA</td>
<td>47</td>
</tr>
<tr>
<td>TBA</td>
<td>92</td>
</tr>
<tr>
<td>TBA</td>
<td>102</td>
</tr>
<tr>
<td>TBA</td>
<td>207</td>
</tr>
<tr>
<td>248</td>
<td>Actor Conditioned Attention Maps for Video Action Detection</td>
</tr>
<tr>
<td>408</td>
<td>Weakly Supervised Gaussian Networks for Action Detection</td>
</tr>
<tr>
<td>463</td>
<td>Weakly Supervised Temporal Action Localization Using Deep Metric Learning</td>
</tr>
<tr>
<td>515</td>
<td>Dynamic Motion Representation for Human Action Recognition</td>
</tr>
<tr>
<td>533</td>
<td>Image to Video Domain Adaptation Using Web Supervision</td>
</tr>
<tr>
<td>534</td>
<td>Stacked Spatio-Temporal Graph Convolutional Networks for Action Segmentation</td>
</tr>
<tr>
<td>559</td>
<td>Global Co-occurrence Feature Learning and Active Coordinate System Conversion for Skeleton-based Action Recognition</td>
</tr>
<tr>
<td>690</td>
<td>Few-Shot Learning of Video Action Recognition Only Based on Video Contents</td>
</tr>
<tr>
<td>811</td>
<td>Action Segmentation with Mixed Temporal Domain Adaptation</td>
</tr>
<tr>
<td>906</td>
<td>Action Graphs: Weakly-supervised Action Localization with Graph Convolution Networks</td>
</tr>
<tr>
<td>981</td>
<td>D3D: Distilled 3D Networks for Video Action Recognition</td>
</tr>
<tr>
<td>1081</td>
<td>Self-Attention Network for Skeleton-based Human Action Recognition</td>
</tr>
<tr>
<td>1096</td>
<td>Long-Short Graph Memory Network for Skeleton-based Action Recognition</td>
</tr>
<tr>
<td>1147</td>
<td>Weakly Supervised Graph Convolutional Neural Network for Human Action Localization</td>
</tr>
<tr>
<td>1167</td>
<td>Temporal Contrastive Pretraining for Video Action Recognition</td>
</tr>
<tr>
<td>197</td>
<td>Fine-Grained Motion Representation For Template-Free Visual Tracking</td>
</tr>
<tr>
<td>610</td>
<td>Adaptive Aggregation of Arbitrary Online Trackers with a Regret Bound</td>
</tr>
<tr>
<td>631</td>
<td>Multiple Object Forecasting: Predicting Future Object Locations in Diverse Environments</td>
</tr>
<tr>
<td>970</td>
<td>Real-time Visual Object Tracking with Natural Language Description</td>
</tr>
<tr>
<td>1060</td>
<td>Inverse Rectification for Efficient Procam Pattern Correspondence</td>
</tr>
<tr>
<td>28</td>
<td>Graph Networks for Multiple Object Tracking</td>
</tr>
<tr>
<td>145</td>
<td>Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison</td>
</tr>
<tr>
<td>792</td>
<td>Video Person Re-identification using Learned Clip Similarity Aggregation</td>
</tr>
<tr>
<td>434</td>
<td>ICface: Interpretable and Controllable Face Reenactment Using GANs</td>
</tr>
</tbody>
</table>
## Oral 2B / Poster 2

**Topic: Object Recognition**

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>Training with Noise Adversarial Network: A Generalization Method for Object Detection on Sonar Image</td>
</tr>
<tr>
<td>41</td>
<td>Active Adversarial Domain Adaptation</td>
</tr>
<tr>
<td>108</td>
<td>Progressive Domain Adaptation for Object Detection</td>
</tr>
<tr>
<td>109</td>
<td>Boosting Standard Classification Architectures Through a Ranking Regularizer</td>
</tr>
<tr>
<td>111</td>
<td>Overlap Sampler for Region-Based Object Detector</td>
</tr>
<tr>
<td>119</td>
<td>A one-and-half stage pedestrian detector</td>
</tr>
<tr>
<td>128</td>
<td>Model-Agnostic Metric for Zero-Shot Learning</td>
</tr>
<tr>
<td>156</td>
<td>Intelligent Image Collection: Building the Optimal Dataset</td>
</tr>
<tr>
<td>162</td>
<td>Internet of Things (IoT) Discovery using Deep Neural Networks</td>
</tr>
<tr>
<td>204</td>
<td>Propose-and-Attend Single Shot Detector</td>
</tr>
<tr>
<td>239</td>
<td>Local Binary Pattern Networks for Character Recognition</td>
</tr>
<tr>
<td>245</td>
<td>Leveraging Filter Correlations for Deep Model Compression</td>
</tr>
<tr>
<td>251</td>
<td>360-Indoor: Towards Learning Real-World Objects in 360° Indoor Equirectangular Images</td>
</tr>
<tr>
<td>269</td>
<td>Regularize, Expand and Compress: NonExpansive Continual Learning</td>
</tr>
<tr>
<td>275</td>
<td>Synthetic Examples Improve Generalization for Rare Classes</td>
</tr>
<tr>
<td>302</td>
<td>CANZSL: Cycle-Consistent Adversarial Networks for Zero-Shot Learning from Natural Language</td>
</tr>
<tr>
<td>307</td>
<td>Accuracy Booster: Performance Boosting using Feature Map Re-calibration</td>
</tr>
<tr>
<td>317</td>
<td>Generating Positive Bounding Boxes for Balanced Training of Object Detectors</td>
</tr>
<tr>
<td>322</td>
<td>Towards Learning Affine-Invariant Representations via Data-Efficient CNNs</td>
</tr>
<tr>
<td>349</td>
<td>Meta-Learning based Filter Pruning for Efficient CNNs</td>
</tr>
<tr>
<td>402</td>
<td>Is Pruning Compression?: Investigating Pruning Via Network Layer Similarity</td>
</tr>
<tr>
<td>404</td>
<td>Transductive Zero-Shot Learning for 3D Point Cloud Classification</td>
</tr>
<tr>
<td>419</td>
<td>Wide Hidden Expansion Layer for Deep Convolutional Neural Networks</td>
</tr>
<tr>
<td>437</td>
<td>Learning from THEODORE: A Synthetic Omnidirectional Top-View Indoor Dataset for Deep Transfer Learning</td>
</tr>
<tr>
<td>508</td>
<td>TKD: Temporal Knowledge Distillation for Active Perception</td>
</tr>
<tr>
<td>273</td>
<td>Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-free Localization</td>
</tr>
<tr>
<td>1067</td>
<td>See the Sound, Hear the Pixels</td>
</tr>
<tr>
<td>1008</td>
<td>How Much Deep Learning does Neural Style Transfer Really Need? An Ablation Study</td>
</tr>
<tr>
<td>362</td>
<td>Kornia: an Open Source Differentiable Computer Vision Library for PyTorch</td>
</tr>
<tr>
<td>986</td>
<td>Multi-Level Representation Learning for Deep Subspace Clustering</td>
</tr>
</tbody>
</table>