Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

Krishna Regmi

M.S., Southern Illinois University Edwardsville, 2015
B.Engg., Tribhuvan University, 2009

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
(COMPUTER SCIENCE)

30 June, 2021, 3:00 P.M.

https://ucf.zoom.us/j/91680522389?
pwd=emQ1U1JWNW52bEllVFNiMTYzU3Ydz09&from=addon

DISSERTATION COMMITTEE

Professor Mubarak Shah, *Chair*, *shah@crcv.ucf.edu*
Professor Scott Branting, *scott.branting@ucf.edu*
Professor Lotzi Bölöni, *ladislau.boloni@ucf.edu*
Professor Yogesh Singh Rawat, *yogesh@crcv.ucf.edu*

DISSERTATION RESEARCH IMPACT

We have developed deep neural network based architectures to generate cross-view images, e.g. aerial image corresponding to the ground-level scene, and also developed a method to subsequently utilize the features of synthesized images to facilitate geo-localization task employing image-based matching. Such localization is useful in vision based navigation in a GPS denied environment, e,g. a moving car can self-localize itself in case of GPS failure. Similarly, given the aerial images of a particular region in the world over time, we can analyze the changes (development or destruction) the region undergoes and study its historical evolution. Additionally, the authenticity of the geo-tagged images/videos uploaded in the social media can be validated, which can help purge the fake information from being circulated in the media.

SELECTED PUBLICATIONS (h-index: 4, total citation: 126)

1. **Cross-view Image Synthesis using Conditional GANs ,** Krishna Regmi and Ali Borji, *in Conference in Computer Vision and Pattern Recognition (CVPR), 2018.*
2. **Cross-view Image Synthesis using Geometry-guided Conditional GANs ,** Krishna Regmi and Ali Borji, *in Computer Vision and Image Understanding (CVIU),* 2019.
3. **From Third Person to First Person: Dataset and Baselines for Synthesis and Retrieval ,** Mohamed Elfeki, Krishna Regmi, Shervin Ardeshir and Ali Borji, *in Conference in Computer Vision and Pattern Recognition Workshops (CVPRW),* 2019.
4. **Bridging the Domain Gap for Ground-to-Aerial Image Matching ,** Krishna Regmi and Mubarak Shah, *in International Conference in Computer Vision (ICCV), 2019.*
5. **Novel View Video Prediction Using a Dual Representation,** Sarah Shiraz, Krishna Regmi, Shruti Vyas, Yogesh Singh Rawat and Mubarak Shah, *in International Conference in Image Processing (ICIP), 2021.*
6. **Video Geo-Localization Employing Geo-Temporal Feature Learning and GPS Trajectory Smoothing ,** Krishna Regmi and Mubarak Shah, *under review in International Conference in Computer Vision (ICCV), 2021.* (submitted)

DISSERTATION

EXPLORING RELATIONSHIPS BETWEEN GROUND AND AERIAL VIEWS BY SYNTHESIS AND MATCHING

Cross-view images, referring to the images taken from aerial and street views, contain drastically differing representations of the same scene of a given location. Due to the differences in the camera viewpoints of ground and aerial images the same semantic concepts in the two viewpoints look very different. Therefore the problem of relating them is very challenging. Thus, it becomes crucial to explore the cross-view relations and learn appropriate representations such that images from these two domains can be associated.

First, we explore supervised approach for **cross view image synthesis** problem to generate realistic images from the target (eg. ground) view, given an image from a source (eg. aerial) view. We solve this problem by utilizing Generative Adversarial Networks (GANs) to synthesize the target images and an auxiliary output, the target view segmentation maps, from source view images. We do so by enforcing the networks to correctly align and orient the different semantics in the scene by jointly penalizing the networks on the quality of target view images and the semantic segmentation maps. We conduct extensive qualitative and quantitative evaluations on Dayton and CVUSA datasets to validate the effectiveness of our methods compared to the baselines.

Next, we propose a novel approach to perform **geometrically-guided cross-view image synthesis**. We leverage the geometrical cues between the aerial and ground images and attempt to preserve the pixels from aerial images to synthesize the ground images. We use homography to transform the aerial images to the street-view and preserve the pixels from the overlapping field of view, followed by inpainting the remaining regions in the ground image. Geometrically transformed images as input ease the network's burden in synthesizing the cross-view images. We conduct extensive evaluations on SVA dataset and demonstrate the superiority of our approach over baselines and previous methods in terms of qualitative and quantitative evaluations.

While cross-view image synthesis is about generating new images from a different viewpoint, we next solve the **cross-view image matching** problem. Here, we find the matching (most similar) image for a query image by computing its feature similarity with the images in the gallery. We propose a novel framework that uses the synthesized images for bridging the domain gap between the images from the two (aerial and ground) viewpoints and helps to learn better features for the cross-view images. These learned features are next employed to solve cross-view geo-localization. Our extensive experiments show that the proposed joint feature learning method outperforms the state-of-the-art methods on CVUSA dataset and with feature fusion, we obtain significant improvements on top-1 and top-10 retrieval accuracies.

Finally, we address the **video geo-localization problem** where we find the matching image for the frames of a video by comparing the frame features with the features of the geo-tagged reference images. We develop a novel method that learns temporally and geographically coherent features for individual frames in the query video by attending to all the frames of the video. We benchmark a new dataset for the problem of video geo-localization that consists of videos from four different regions of the USA. We conduct extensive evaluations to validate that the proposed approach performs better compared to methods that learn image features independently and our method generalizes well to different regions of the USA.

Krishna Regmi

1985	Born in Chitwan, Nepal
2009	B.Engg., Tribhuvan University, Kathmandu, Nepal
2015	M.S., Southern Illinois University Edwardsville ,Illinois
2016-2021	Ph.D., University of Central Florida, Orlando, FL
2016-2017	ORC Doctoral Fellowship
2019	Research Scientist Intern, Netflix, Los Gatos, CA