Dataset Labelling

Dataset preparation has the following stages:

1. Collection of satellite data products of both optical and microwave sensors which overlap in a short date range if not same date of pass.
2. Carefully co-register the SAR and optical pairs.
3. Clipping the intersecting area
4. Getting both the cropped image's resolutions to match
5. Tiling and storing the data with appropriate file naming convention.

An initial test run has been done on the following pair by the author to derive the methodology

a. Optical image : C2E data over a city in andhra pradesh (PID: 21398721)
b. SAR image: RISAT-2B-R1 that contains the city in it's image extents (PID:193128811)

Detailed Steps for Dataset Preparation:

1. Collection of datasets:
    - Initially a catalogue of SAR imagery has been prepared collecting the geolocation, date of pass, path, row, image file location, etc and a database has been created.
    - Manually, few samples are selected with urban regions falling in the extents.
    - Based on the samples selected and the corresponding data of pass and image extents, optical images (preferably C2E) should be fetched from bhoonidhi-paid portal.
    - Each collected pair to be stored in separate folders in the following convention.
      /raw/<optical PID>_<SAR PID>/optical/.
      /raw/<optical PID>_<SAR PID>/sar/.
2. Co-registering using QGIS: (inputs: /raw/<optical PID>_<SAR PID>/optical/BAND.vrt, /raw/<optical PID>_<SAR PID>/sar/<SAR PID>_VV.tif)
    - After opening QGIS version 3 and above, load the SAR image VV polarisation (since popular and more oftenly used) on the main map.
    - Open the optical data in the "Georeferencer" utility. It can be found in the toolbar under Raster->Georeferencer.
    - Start adding GCP points in the geo-referencer tool, after finding a good geo-feature. Once a point is found and clicked, the corresponding GCP shall be picked from the main map canvas (i.e SAR image on the main map).
    - It is to be ensured that at least 5 points from each corner of the optical image are tagged to avoid internal distortion.
    - Once, the GCPs are set, the output settings are to be configured. The projection is to be set the same as the SAR image (which can be found by clicking properties). The image is to be saved in the folder.
      /geolinked/<optical PID>_<SAR PID>
    - Even the SAR image is to be moved to the above folder.

3. Clipping the intersection: (inputs: /geolinked/<optical PID>_<SAR PID>/optical.tif, /geolinked/<optical PID>_<SAR PID>/sar.tif)
    - A shapefile is to be created Layer->create layer->shapefile. And the type to be set as polygon.
    - Then edit and draw a polygon following the intersection region of the SAR and optical.
    - Now, both the images to be clipped Raster->Extraction->Clip by mask layer
    - Save the both images in the following format
      /geolinked/<optical PID>_<SAR PID>/optical_clipped.tif
      /geolinked/<optical PID>_<SAR PID>/sar_clipped.tif
4. Match the resolutions (inputs: /geolinked/<optical PID>_<SAR PID>/optical_clipped.tif, /geolinked/<optical PID>_<SAR PID>/sar_clipped.tif)
    - Check the resolution of the optical image (rows and columns)
    - Then Raster->Raster Calculator
    - On the left panel, select the sar_clipped.tif and double click
    - Similarly on the right panel, change the resolution to the same resolution as the optical_clipped.tif.
    - Save the image in the following format
      /geolinked/<optical PID>_<SAR PID>/sar_final.tif
    - Rename the optical_clipped.tif as optical_final.tif
5. Tiling
    - Tiling can be done using the python script attached along with the document.
    - All the generated tiles are stored in the following format
      /tiles/sar/sar_<optical_PID>_<SAR PID>_<i>_<j>.tif
      /tiles/optical/opt_<optical_PID>_<SAR PID>_<i>_<j>.tif

➔ A catalogue of SAR data is collected and kept in /maintenance/WPDSD_IMP/TDP/Catalogue.csv

➔ Based on the dop and geo location, corresponding C2E images are to be extracted.

Agenda of meeting to be scheduled on February 2021:

1. A trial preparation of a dataset.
2. Curation of urban data from the catalogue.
3. Collection of C2E data from paid-bhoonidhi or UOPS (whichever is working at the time).
4. Work distribution and progress tracking strategies.
5. Discussion of the last meeting's actions: Literature survey, previous existing methods, system setup for simulations, preliminary simulation results discussion.
6. Plan of action