LONDON imaging MEETING 2021
20-22 SEPTEMBER
IMAGING FOR DEEP LEARNING

Joint Unsupervised Infrared-RGB Video Registration and Fusion

Imad Eddine Marouf[1,2], Lucas Buras[2]
Hakki Can Karaimer[2,3], Sabine Süsstrunk[2]
[1]IP Paris [2]IC, EPFL [3]Advanced Micro Devices (AMD)

Introduction

- In 2011, 70,090 firefighters in the U.S. alone were injured in the line of duty with, 61 deaths.

- Poor visibility might lead to influences of human behavior such as redirection of movement and their initial response speed.

Solution: Combined usage of infrared (IR) and RGB Cameras.

- ❑ IR cameras to aid in seeing through smokes and detect the variation of temperature in the surroundings.
- ❑ While IR cameras' response is related to the temperature within the captured frame, RGB cameras' response is related to human perception.

Background

Joint video registration and fusion of IR-RGB video pairs consist of:

Video registration that aligns videos of the same scene, and

video fusion that brings all the essential information from the two video modalities to a single video.

Proposed Method

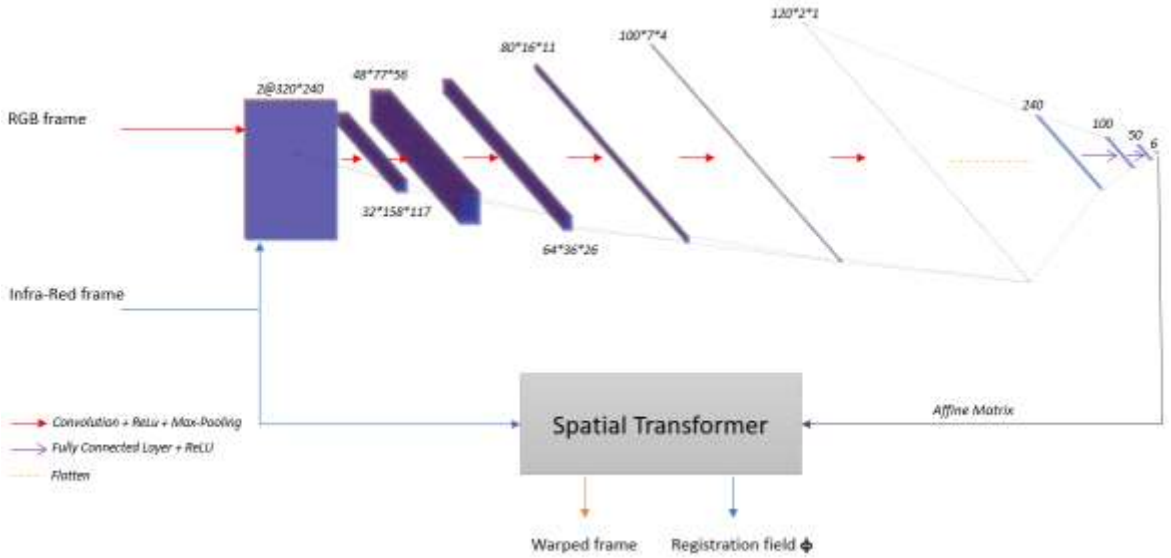
Our pipeline consists of:

Semantic segmentation that generates masks for objects using *Mask-RCNN*,

registration that creates well-matched masks using the *affine network*,

fusion to fuse registered frame with RGB using *Zero-Learning Fusion* model, and

alpha blending to restore the original RGB image's colors.



Our proposed *affine network* is a spatial transformation network that results in an affine matrix applied on the IR frame.

LONDON imaging MEETING 2021

Results

- Applying image registration with VoxelMorph/affine networks directly to video pairs leads to poor results:
  - Shape deformations and misaligned images.
- Applying semantic segmentation for humans prior to image registration provides a good basis to have well-matched IR-RGB pairs.

IR-RGB registration with the VoxelMorph architecture and the affine network. When VoxelMorph (2nd and 3rd row) is trained with $\lambda = 0.6$ and $\lambda = 0.1$, it poorly deforms the shape of the objects in the image. When compared visually, the affine network (1st row) achieves better results.

Conclusion

- A framework to assist firefighters in performing their missions in extreme visibility conditions
- Accurate segmentation of persons in IR and RGB image pairs leads to well-matched IR-RGB video pairs.

LONDON imaging MEETING 2021