\iclrconference

MULTI-MODAL SENSOR REGISTRATION FOR VEHICLE PERCEPTION VIA DEEP NEURAL NETWORKS

Michael Giering, Kishore Reddy, Vivek Venugopalan
Decision Support
Machine Intelligence Group United Technologies Research Center
E. Hartford, CT 06060, USA
{gierinmj, reddyk, venogvi}@utrc.utc.com

ABSTRACT

When performing multi-modal fusion to perform an analytic task, spatio-temporal registration of the incoming signals is often a prerequisit to analyzing the fused data and critical to the stability of the analysis. Lidar-Video systems like on those many driverless cars are a common example of where keeping the Lidar and video channels registered to common physical features is important. We develop a deep learning method that takes multiple channels of heterogeneous data to detect the misalignment of the Lidar-video inputs. A number of variations were tested on the Ford LV driving test data set with minimal tuning of the deep conv nets parameters.

1 MOTIVATION

Navigation and situational awareness of optionally manned vehicles requires the integration of multiple sensing modalities such as LIDAR and video, but could just as easily be extended to other modalities including Radar, SWIR and GPS. Spatio-temporal registration of information from multimodal sensors is technically challenging in its own right. For many tasks such as pedestrian and object detection tasks that make use of multiple sensors, decision support methods rest on the assumption of proper registration. Most approaches [][] in LIDAR-video for instance, build separate vision and lidar feature extraction methods and then try to identify common anchor points in both. Generating a single feature set on Lidar, Video and optical flow, enables the system to to capture mutual information among modalities more efficiently. The ability to dynamically register information from the available data channels for perception related tasks can alleviate the need for anchor points *between* sensor modalities. We see auto-registration as a prerequisit need for operating on multi-modal information with confidence.

Deep neural networks lend themselves in a seamless manner for data fusion on time series data. It has been shown [Ng multimodal] for some problems that features generated on the fused information [] can provide insight that neither input alone can. In effect the ML version of, ”the whole is greater than the sum of it’s parts”.

Speed constraints of real time navigation also constrain model selection. The trained nnets easily run within the real-time constraints of common frame rates and lidar data collection. *Kishore, let me know if this isn’t true for optical flow*

From an applied research perspective, it is possible to create such systems with far less overhead. The need for domain experts and hand-crafted feature design are lessened, thereby allowing more rapid prototyping and testing.

The generalization of autoregistration across multiple assets is clearly a path to be explored.

By including optical flow as input channels, we imbue the nnet with information on the dynamics observed across time steps.

2 PREVIOUS WORK

Need some references to define the state of the art

3 PROBLEM STATEMENT

These arrays were subdivided into p x p x C patches at a prescribed stride. For any experiment we can denote the preprocessing parameters

- R,G,B — Frame color channels.
- U,V — optical flow channels.
- L — lidar depth channel.
- C — number of input channels.
- p — patch size.
- s — stride.

For a given frame of size 800 x h there are approximately n= (800 x h)/s patches (exact number?). The training and test sets had X and Y frames respectively, therefore the entire data set consists of N = n x X inputs of the patch-size dimension.

Preprocessing is repeated O times, where O is the number of offset classes. For this work we used two setups. A 5 class, linearly distributed set of offsets and a 9 class eliptically distributed set of offsets. (see figure x) For each offset class, **Kishore explain how you generated the data.**

In order to accurately detect misalignment in the LV sensor data, we've assumed there needs to be a lower bound on the amount of information present in each channel. For this data set, L was the only channel with regions of low information. A preprocess step was to eliminate all patches corresponding to L data with variance ¡x. This leads to the elimination of the majority of foreground patches in the data set, reducing the size of the training set by **z pct KISHORE**

4 MODEL DESCRIPTION

need to describe the parameters post-processing,classification metric for each patch,a table with common params for the experiments would help,voting scheme

The model consists of a 4-layer **?** CNN classifier *see image of network* that estimates the offset between the LV inputs at each time step. For each patch within a timestep, there are O variants with the LVF inputs offset by the predetermined amounts. The CNN outputs to a softmax layer, thereby providing an offset classification value for each patch of the frame. figure x: In the 5 class example we color each patch of the frame with a color corresponding to the predicted class.

For each frame a simple voting scheme is used to aggregate the patch level offset predictions to frame level predictions. A sample histogram of the patch level predictions is show in figure x.

4.1 OPTICAL FLOW

kishore, please discuss the motivation to include dynamics, how we performed it and how we'd need to do it if running in real time. this is where we can point ot proof that it improves prediction.

5 EXPERIMENTS AND POST-PROCESSING

Need a complete list of the experiments run images to visualize the frame level results please place any confusion matrices and your comments on what you think the results say. feel free to suggest any tables or other visuals to include.

5.1 5 CLASS TESTS

In our initial tests, the linearly distributed set of 5 offsets of the LV data were performed. Table 1 lists the inputs and CNN parameters explored ranked in the order of increasing accuracy **(define accuracy and other cm metrics), include training vs test error and conf mats if room allows**.

As can be seen ...

5.2 9 CLASS TESTS

The subsequent tests were designed to understand whether the simple linear displacement model of the 5-class test could be generalized to a model capable of discriminating multiple directions and displacement magnitude. To acheive this 8 positions were chosen on an ellipse along with it's center **describe the parabola**. LV was offset in a manner similar to the 5 class test. Nine training and test sets were generated and an identicle patch level CNN was constructed differing only in the 9 class softmax output layer.

Table 2 lists the inputs and CNN parameters explored ranked in the order of increasing accuracy **(define accuracy and other cm metrics), include training vs test error and conf mats if room allows**.

Discussion: what results confirmed expectations or surprised us (grey scale). Can we confidently say optical flow improves prediction.

6 CONCLUSION AND FUTURE WORK

We did it. We're great.

future: implement a method that doesn't require ground truth and also generalizes easily to a wide array of sensors. Test it on data collected from airborne platforms that are noisier and have more degrees of freedom.

7 REFERENCES

populate the papers to be cited in the folder and if possible the bib file

8 GENERAL FORMATTING INSTRUCTIONS

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in small caps and left-aligned. All pages should start at 1 inch (6 picas) from the top of the page.

Authors' names are set in boldface, and each name is placed above its corresponding address. The lead author's name is to be listed first, and the co-authors' names are set to follow. Authors sharing the same address can be on the same line.

Please pay special attention to the instructions in section **??** regarding figures, tables, acknowledgments, and references.

9 HEADINGS: FIRST LEVEL

First level headings are in small caps, flush left and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

9.1 HEADINGS: SECOND LEVEL

Second level headings are in small caps, flush left and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

9.1.1 HEADINGS: THIRD LEVEL

Third level headings are in small caps, flush left and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

10 Citations, figures, tables, references

These instructions apply to everyone, regardless of the formatter being used.

10.1 Citations within the text

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis (as in "See **?** for more information."). Otherwise, the citation should be in parenthesis (as in "Deep learning shows promise to make progress towards AI (**?**).").

The corresponding references are to be listed in alphabetical order of authors, in the References section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

10.2 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

10.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

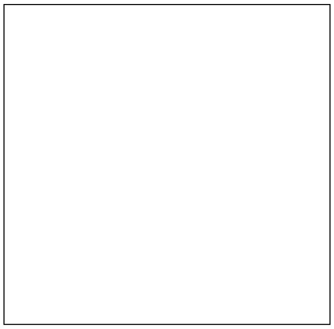

Figure 1: Sample figure caption.

10.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table **??**.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

[1] Sample of the first footnote

[2] Sample of the second footnote

Table 1: Sample table title

PART	**DESCRIPTION**
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

11 FINAL INSTRUCTIONS

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

12 PREPARING POSTSCRIPT OR PDF FILES

Please prepare PostScript or PDF files with paper size “US Letter”, and not, for example, “A4”. The -t letter option on dvips will produce US Letter files.

Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.

Otherwise, please generate your PostScript and PDF files with the following commands:

```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

12.1 MARGINS IN LATEX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the graphicx package. Always specify the figure width as a multiple of the line width as in the example below using .eps graphics

```
   \usepackage[dvips]{graphicx} ...
   \includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
   \usepackage[pdftex]{graphicx} ...
   \includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (`http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command.

ACKNOWLEDGMENTS

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.