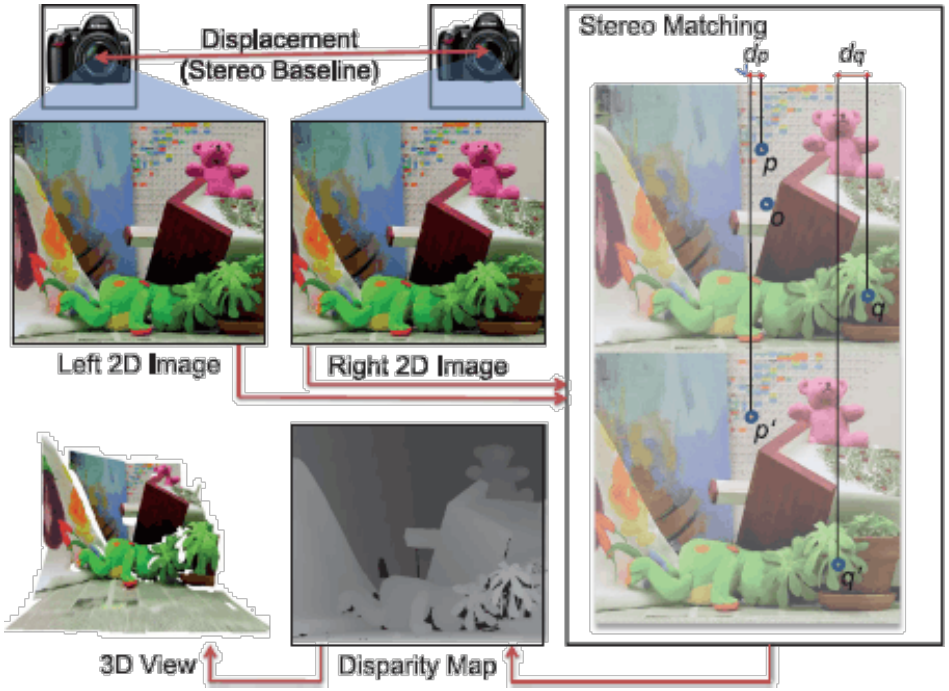
A Machine Learning Approach to Solving the Correspondence Problem in Computer Vision

Ronald Cruz, Paul Fisher

New York University

Foundations of Machine Learning 2017

The Problem



https://doi.org/10.1007/978-1-4471-5520-1_6 [1]

- Pair of stereo images, slightly differing due to camera movement
- Matching pixels between the images
- Disparity inversely proportional to distance from viewer

Outline

- Correspondence as binary classification
- Data set
- Feature selection
- Algorithm used
- Results

Correspondence as Binary Classification

- Local Methods - match windows around pixel (Winner takes all)

$$d_p = \underset{0 \leq d \leq d_{max}}{\operatorname{argmin}} \sum_{q \in W_q} w(p,q) \cdot c(q, q-d)^1$$

- $w(p,q)$ is a measure of how similar pixels p and q are
- Assumption: spatially close points exhibit similar disparity
- Goal: Learn a function w
  - Learning a kernel?

[1] https://doi.org/10.1007/978-1-4471-5520-1_6 [1]

Data

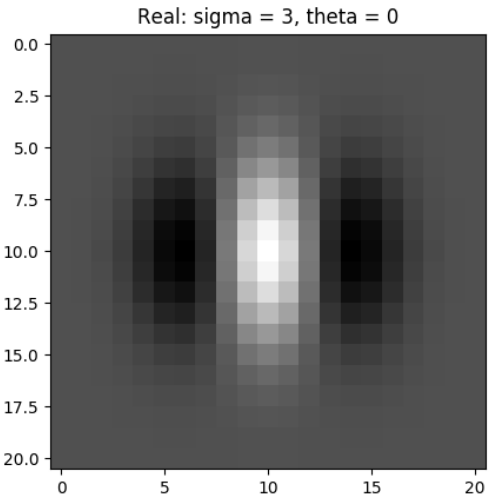
- Dataset: 2014 Middlebury Stereo
- Total of 33 high resolution images
  - Utilized a subset of 10 training and testing sets.

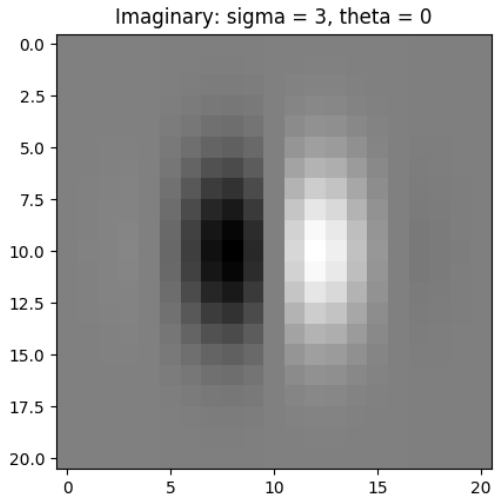

http://vision.middlebury.edu/stereo/data/2014/

Feature Selection

- Sample represents a pair of pixels - one from left image and one from the right
- pixel-wise differences in a 25 $\times$ 25 window surrounding selected pixels.
- Convolve with 2-d Morlet Wavelet[2]:

$$\psi_{\sigma,\theta}(\boldsymbol{u}) = \frac{C_1}{\sigma}\left(e^{i\frac{\pi}{2\pi}(\boldsymbol{u}\cdot e_\theta)} - C_2\right)e^{-\frac{\boldsymbol{u}^2}{2\sigma^2}}$$





[2] Davi Geiger CSCI-GA.3033-012 "Haar Basis wavelets and Morlet Wavelets"

Algorithm Used

- Linear Support Vector Machine with SGD for training
  - Scales with large number of samples
  - Can be trained on-line
- Trained on selected portions of images
- Line search to select best matching for pixel (based on confidence)

Example Results

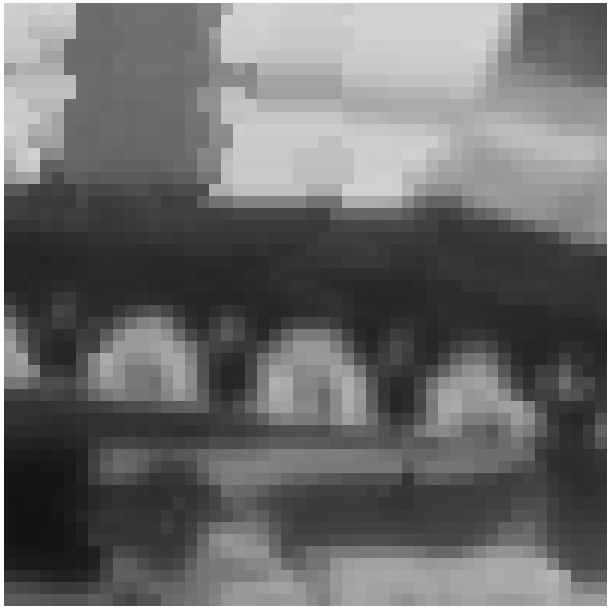
A small portion of the actual left image in gray

The corresponding pixels in the right image, as predicted by our model

- Predicted disparity within 1 pixel: 44.8%
- Within 2 pixels : 68.1%
- Within 5 pixels : 88.6%
- Mean square error: 8.36 pixels
- Accuracy highly dependent on features of the image portion

Example Results

A small portion of the actual left image in gray



The corresponding pixels in the right image, as predicted by our model

- Predicted disparity within 1 pixel: 65.4%
- Within 2 pixels : 87.0%
- Within 5 pixels : 95.0%
- Mean square error: 2.24 pixels
- Accuracy highly dependent on features of the image portion

Discussion

- Sensitivity: Occlusion introduces significant noise
- Difficulty of matching large areas uniform in color
- Paths to explore in future:
  - Different wavelets
  - Smaller features space while utilizing Kernel function
  - Neural Networks

References

Michael Bleyer and Christian Breiteneder. *Stereo Matching—State-of-the-Art and Research Challenges*, pages 143–179. Springer London, London, 2013.

A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann. Local stereo matching using geodesic support weights. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2093–2096, Nov 2009.

Kuk-Jin Yoon and In-So Kweon. Locally adaptive support-weight approach for visual correspondence search. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 924–931 vol. 2, June 2005.