A General Framework for Fast 3D Object Detection and Localization Using an Uncalibrated Camera

Authors: Andres Solis Montero, Jochen Lang, and Robert Laganiere

University of Ottawa
School of Electrical Engineering and Computer Science
VIVA Lab

Problem & Objectives

Problem

- Visual real-time 3D object detection
- Mobile camera independent of camera optics and mobile object
- Memory constraints and no 3D model of the object

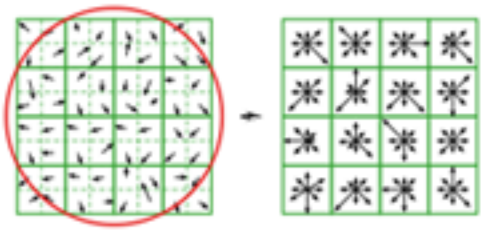
Objectives

Real-time single camera video-based application:

- detects moving objects from an uncalibrated mobile camera
- has small memory footprint
- is invariant to viewpoint changes, robust to noise and image illumation changes
- accounts for occlusions and cluttered environments

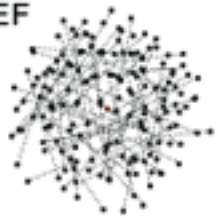
Related Work

SIFT

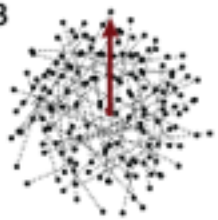


SIFT, D.G. Lowe. 2004
SURF, H. Bay et al. 2008
3D object recog. Rothganger et al. 2004

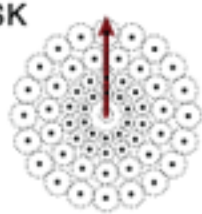
BRIEF



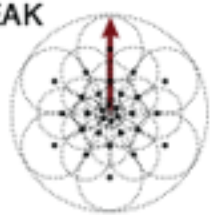
ORB

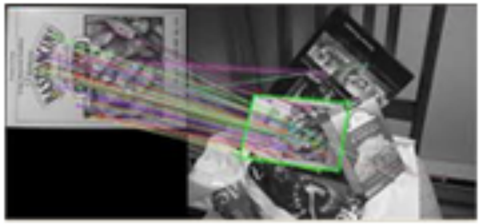


BRISK

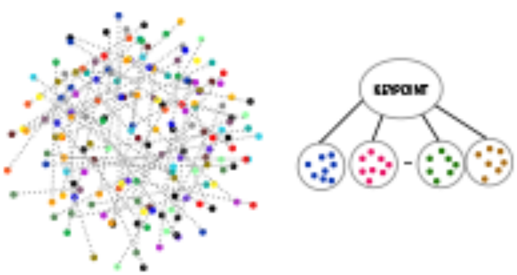


FREAK





FERNS

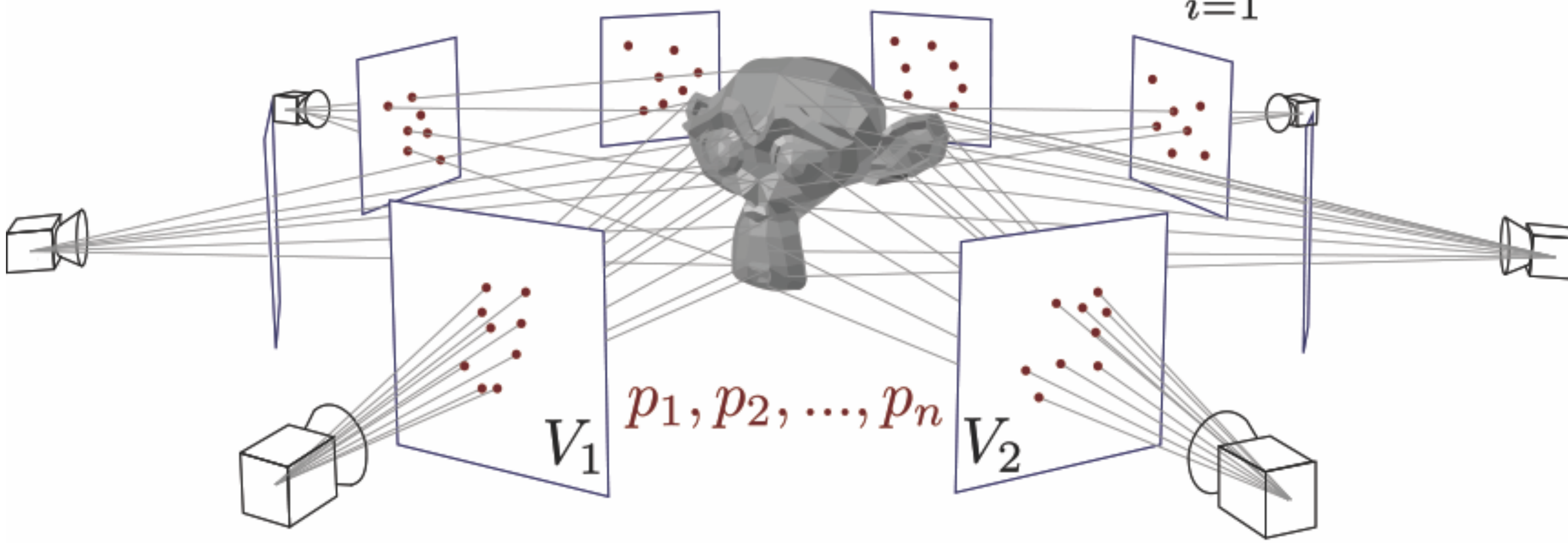


BRIEF, M. Calonder et al. 2010
ORB, E. Rublee et al. 2011
BRISK, S. Leutenegger et al. 2011
FREAK, P. Vandergheynst et al 2012
Target location, Taylor & Drummond. 2009
Ferns, M. Ozuysal et al 2010

VIVA uOttawa

Details

$$\arg\max_j P(V_j|F_{j1}, F_{j2}, ..., F_{jn}) = \arg\max_j \prod_{i=1}^{n} P(F_{ji}|V_j)$$



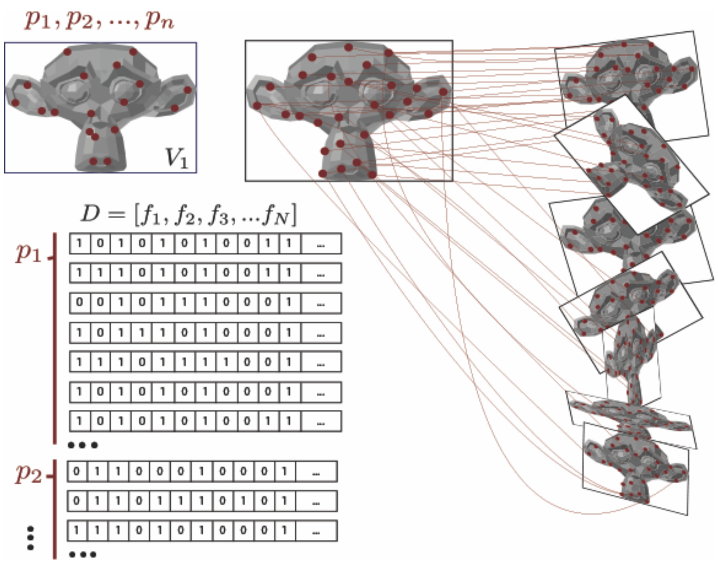
$$V = \{V_j : 1 < j < m\}$$

$$V_j = \{p_1, p_2, ..., p_n\}$$

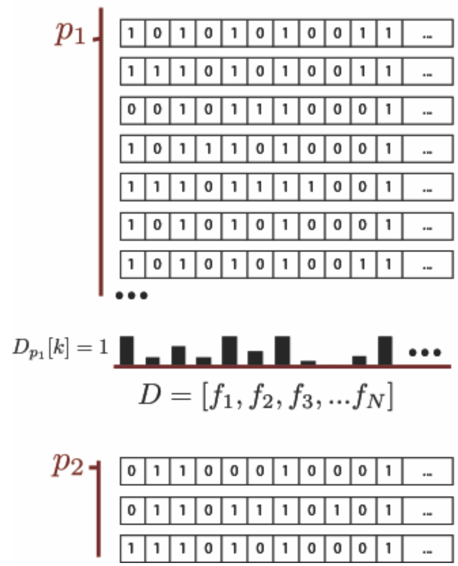
$$F_{ji} = \begin{cases} 1 & p_i \text{ belongs to } V_j \\ 0 & \text{otherwise} \end{cases}$$

$$\prod_{i=1}^{i=n} P(F_{ji}|V_j) = B(n_b, n, P_0)$$

Details

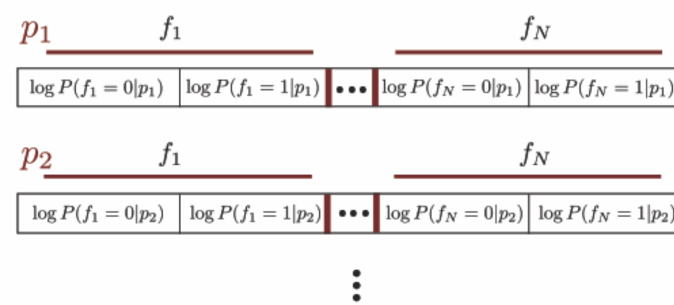


$$\arg\max_i \log P(p_i|f_1, f_2, .., f_N) = \arg\max_i \sum_{k=1}^{N} \log P(f_k|p_i)$$



$$P(f_k|p_i) = \frac{|D_{p_i}[k] = 1| + Nc}{D_{p_i} + K * Nc}$$

$$f_i = \begin{cases} 1 & I_P <_\beta I_{P+1} \\ 0 & otherwise \end{cases}$$



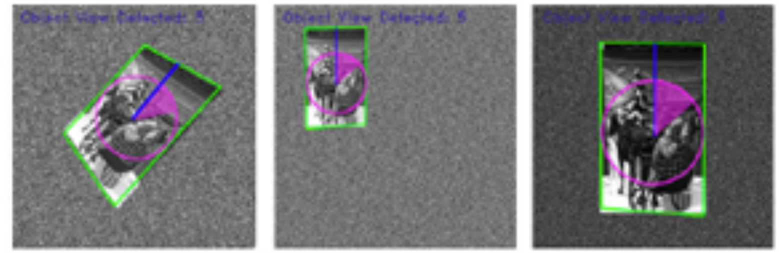
Experimental Results

Experimental Results

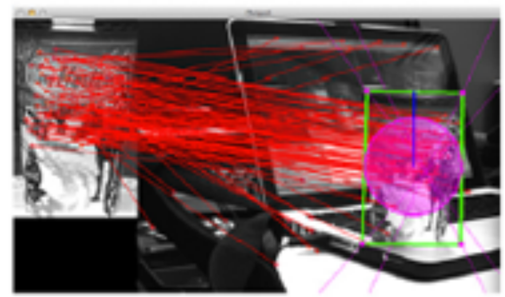
Detection Rates under transformations

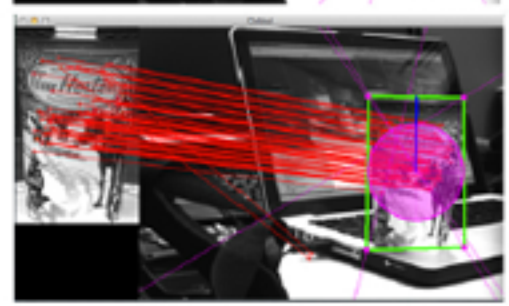
Algorithm	Rotation*	Scale×	Illumination+
BRIEF	9%	43%	62%
NBCD + BRIEF	15%	62%	64%
ORB	75%	74%	89%
NBCD + ORB	92%	91%	90%
BRISK	79%	75%	82%
NBCD + BRISK	94%	92%	84%
FREAK	69%	70%	79%
NBCD + FREAK	84%	89%	81%
Ferns	14%	46%	61%
Ferns*	89%	88%	68%

* roll = $[-180°, 180°]$, pitch and yaw = $[-70°, 70°]$
× scale = $[0.2, 2]$
+ $\alpha \times I(x, y) + \beta, \alpha = [.3, 3], \beta = [-100, 100]$



Our classifier vs Hamming distance





Stats: Ferns moused pad example

Algorithm	Fps	Detection	Memory	Time
Ferns	25.1	**87.4%**	16Mb	348s
NBCD + ORB	**26.5**	86.9%	**80Kb**	**24s**

Contributions & Conclusions

- Framework for real-time 3D object detection using a single, mobile and uncalibrated camera
- Combine binary descriptors with Naïve Bayes classifiers for feature classification and matching
- The new classifier exploits the specific structure of binary descriptors to increase feature matching while conserving descriptor properties
- Small memory footprint due to efficiently encoded features
- Learning time is reduced because invariant features and descriptors
- Improved indexing scheme to speed up keypoing matching