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

Authors: Andrés Solís Montero, Jochen Lang, and Robert Laganière

uOttawa

IEEE Winter Conference on Applications of Computer Vision

1. Introduction

We developed a feature based framework composed of two Naive Bayes Classifiers and a Kalman filter to recognize objects. Our solution exploits binary descriptor properties to create a computational efficient solution with a small memory footprint. Our algorithm recognizes moving objects from a moving camera.

2. The 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
- Solutions must account for occlusions, cluttered environments, viewpoint changes, robust to noise and illumination changes.

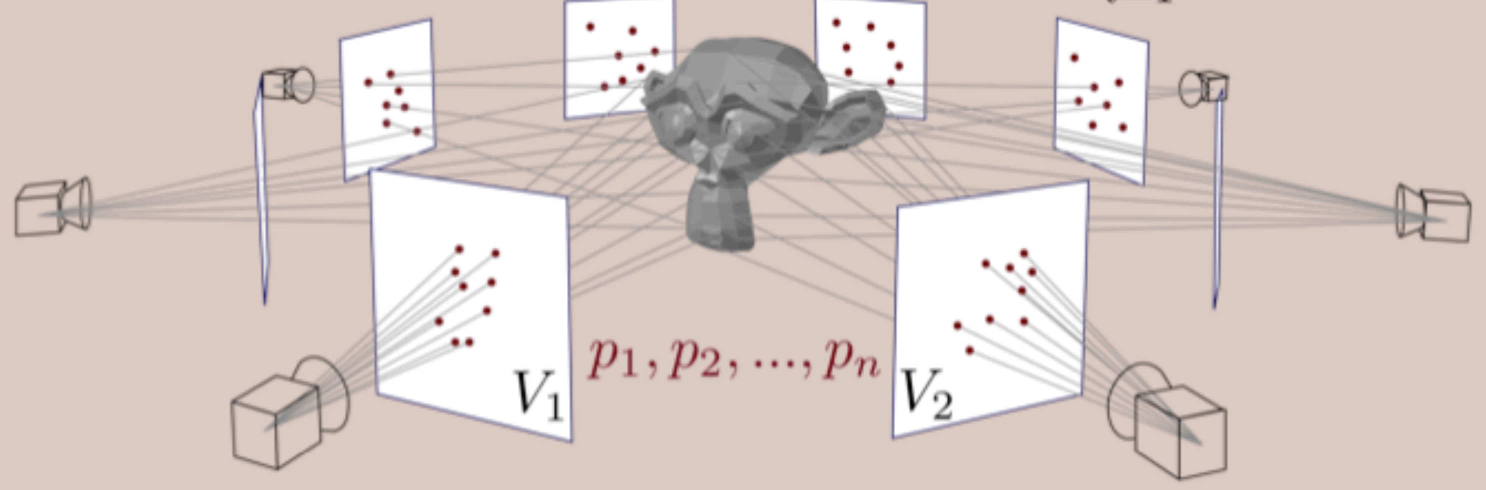
3. Contributions

- 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

4. Naïve Bayes Classifier for View Matching

We learn the 3D object from a set of images (i.e., views) and solve the detection problem by selecting the view (V) with maximum likelihood given a set of keypoint observations.

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

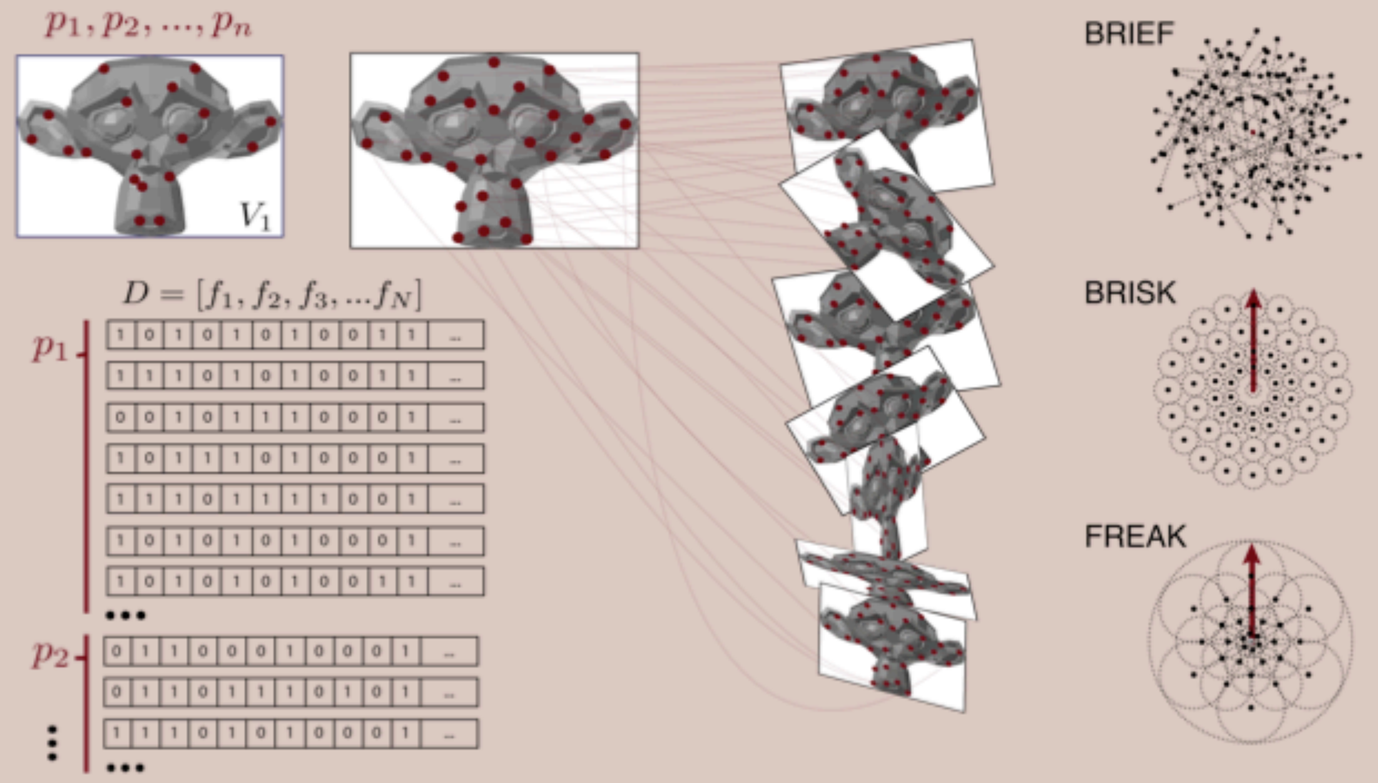


$$V = \{V_j : 1 < j < m\} \qquad F_{ji} = \begin{cases} 1 & p_i \text{ belongs to } V_j \\ 0 & \text{otherwise} \end{cases}$$

$$V_j = \{p_1, p_2, ..., p_n\} \qquad \prod_{i=1}^{i=n} P(F_{ji}|V_j) = B(n_b, n, P_0)$$

5. Learning the Most Stable Keypoints

We synthetically transform each view and select the prominent keypoints with their binary descriptors. Each keypoint is back-projected to the initial view and we select the most repeated ones. The descriptors associated to each stable keypoint are merged in our Naïve Bayes Classifier for keypoint matching.



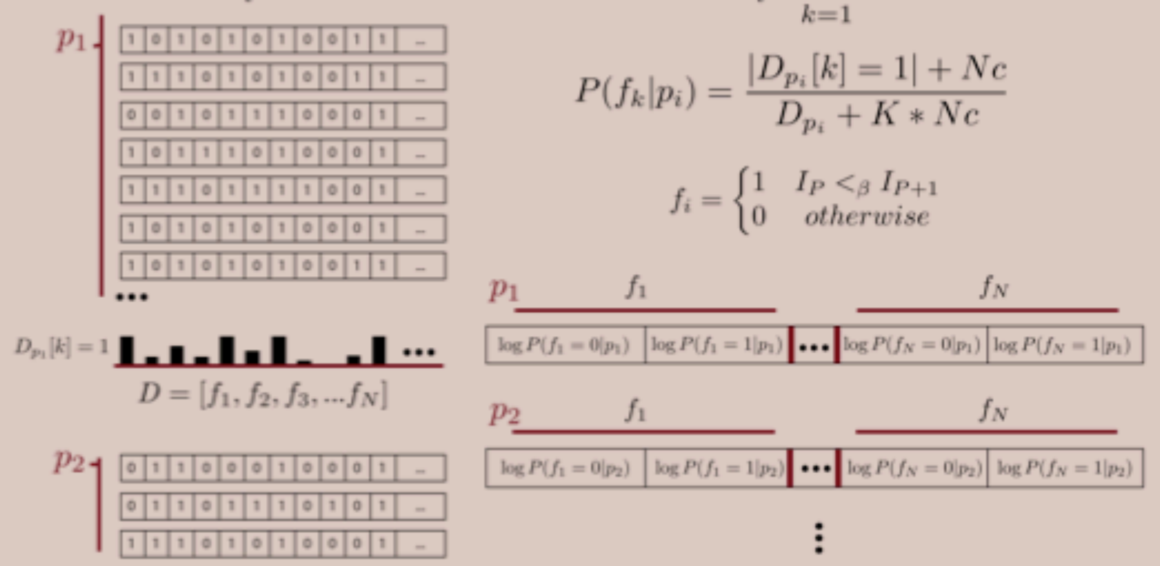
6. Naïve Bayes Classifier for Keypoint Matching

We assume complete independency of each bit from the binary descriptor and encode the probability of each bit to belong to each class (i.e., keypoint).

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

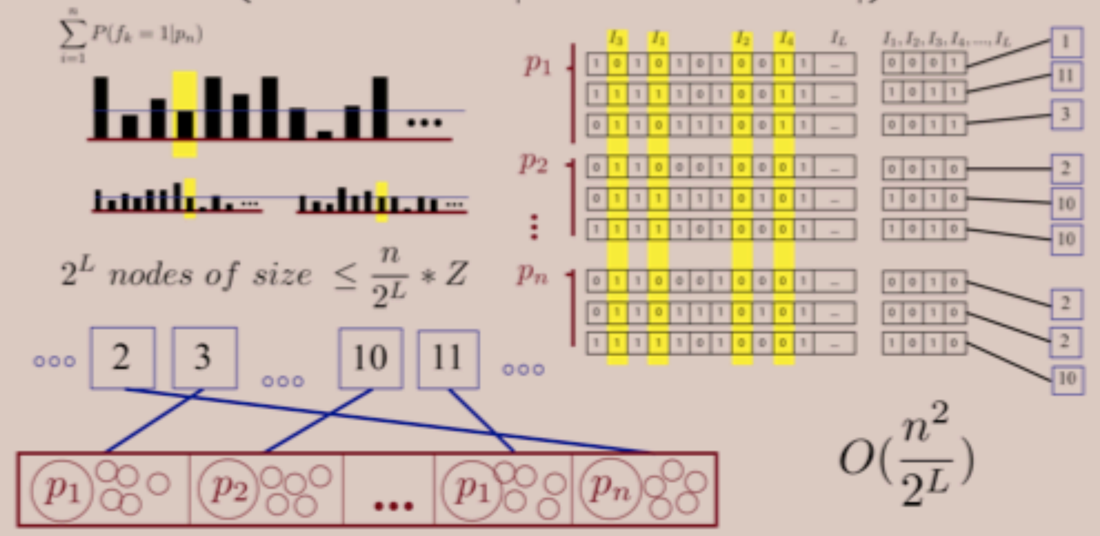
$$P(f_k|p_i) = \frac{|D_{p_i}[k] = 1| + Nc}{D_{p_i} + K * Nc}$$

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

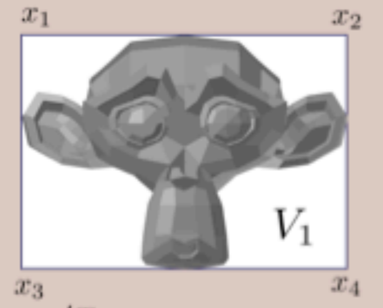


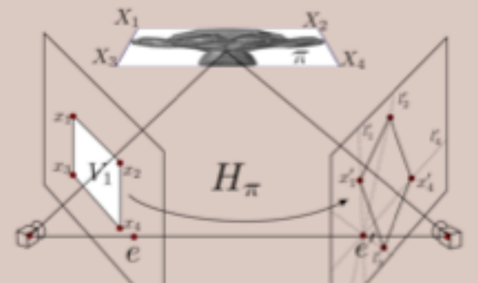
Keypoints are indexed to speed up the matching; combining the bits position that better separate all the descriptor belonging to the same keypoint generates the index.

$$I = \left\{ I_L \middle| L < N, \arg\min_L \left| 0.5 - \frac{\sum_1^n P(f_L = 1|p_n)}{n} \right| \right\}$$

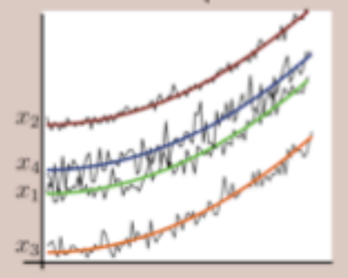


Features from the video are classified and we compute the closest view to the object from our database. The object is projected in the video using the fundamental matrix and the epipolar constraint.





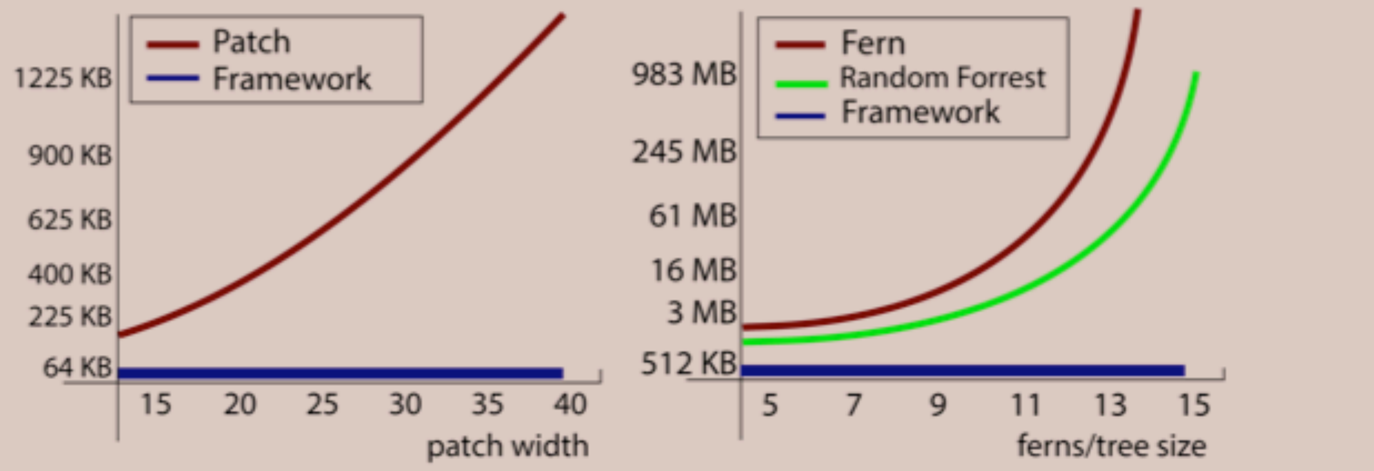
$$x_i'^T F x_i = 0, \quad l_i' = F x_i$$
$$x_i' = H_\pi x_i$$
$$l_i' = [e']_\times x_i' = [e']_\times H_\pi x_i = F x_i$$
$$F = [e']_\times H_\pi$$
$$H_\pi = [e']_\times F$$



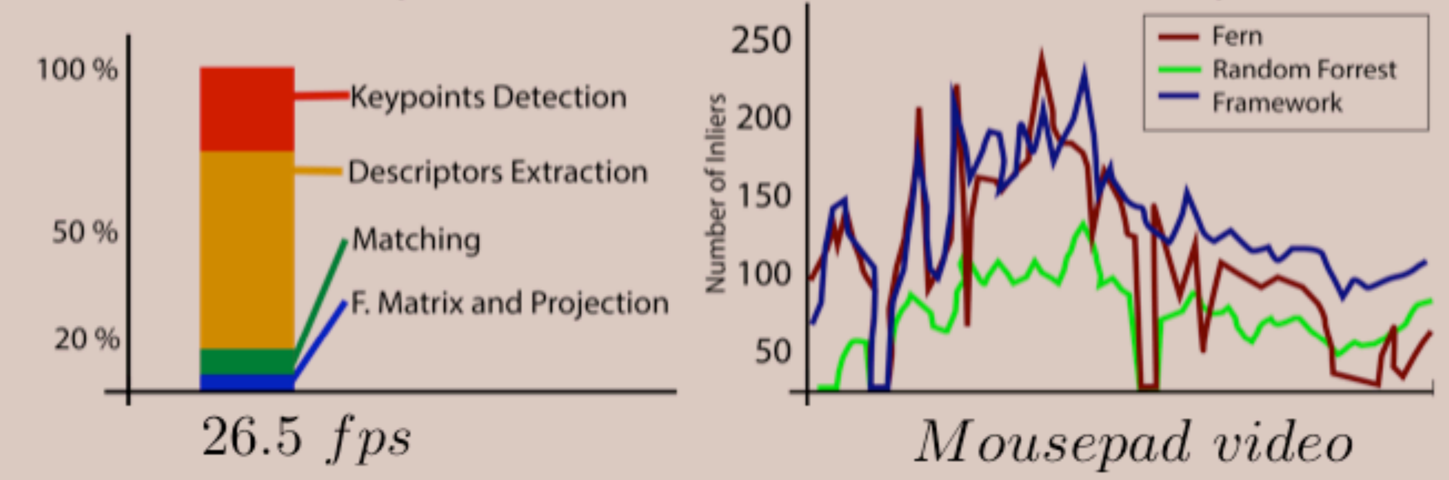
Corner Traking with 2D Kalman Filter

7. Results

Memory usage for 1000 keypoints (offline & online)

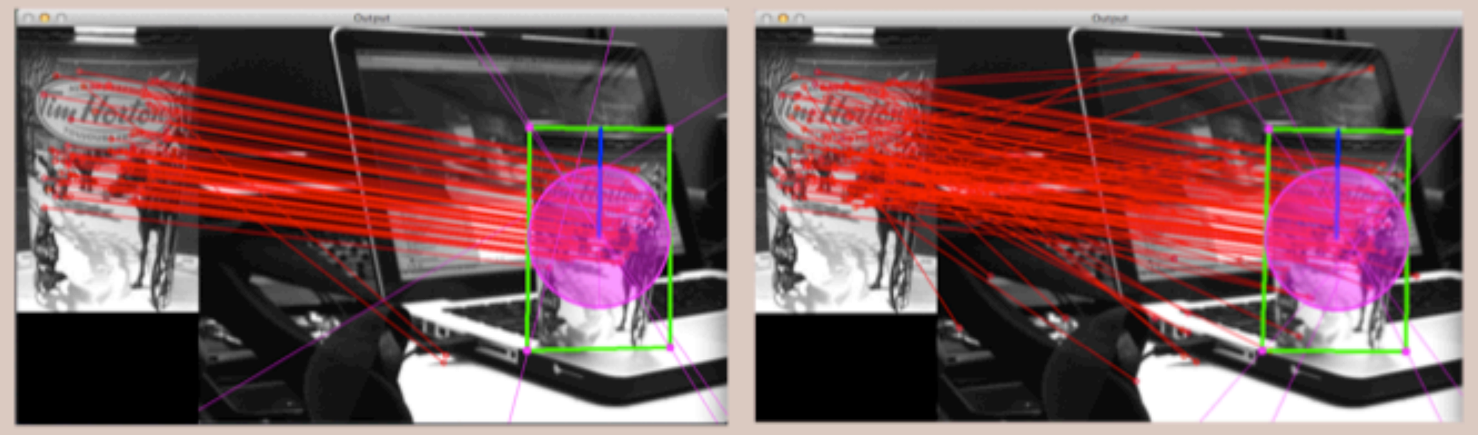


Performance (Execution Time & Detection)



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

Classifier vs Hamming Distance



8. Conclusions

The presented framework archives high recognition rates for textured objects in real time (23 fps). The implementation is capable of detecting 3D moving objects from different viewpoints even when they are partially occluded from view. The experimental results confirm the effectiveness of the proposed solution.