Visual Semantics from an Accelerated Model Extraction and Construction

Leonardo Chang[1,2], Miguel Arias-Estrada[1], José Hernández-Palancar[2], L. Enrique Sucar[1] and Liliana Barbosa-Santillán[3]

[1] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)
Luis Enrique Erro No.1, C.P. 72840, Tonantzintla, Puebla, Mexico.
{lchang, ariasmo, esucar}@ccc.inaoep.mx

[2] Advanced Technologies Application Center (CENATAV)
7ᵐA No. 21406, Playa, C.P. 12200, La Habana, Cuba.
{lchang, jpalancar}@cenatav.co.cu

[3] University of Guadalajara (UDG), Mexico
Periférico Nte 799, Los Belenes, C.P. 45100, Jalisco, Mexico.
ibarbosa@cucea.udg.mx

1. Abstract

• Image representation is a critical step in computer vision.
• However, it remains as one of the most challenging topics, partly because of the lack of su¨cien tly discriminative and robust representations, and the high computational cost of state-of-the-art methods.
• In our work, we propose an image representation method that integrates low/middle-level features extracted from images with high-level cognitive representations, which is accelerated with the use of GPUs.

2. Motivation

• The accuracy and quality of the object recognition process could be improved by introducing high-level semantic knowledge representations.

• Additionally, it would contribute to a cognitive object recognition and interrelation with its scenario.

3. Discussion

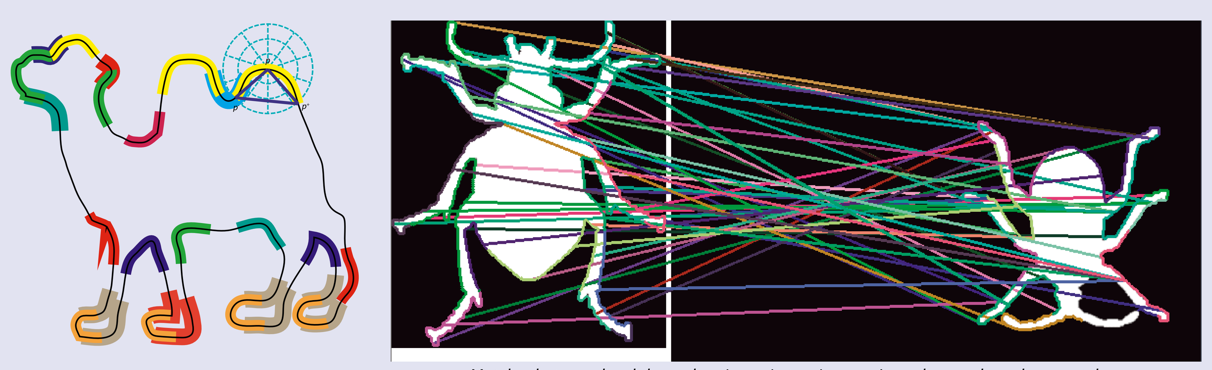
• Integrating low/middle-level features extracted from images with high-level cognitive representations shall contribute to a less structured representation of objects, providing a more accurate representation of real world objects.

• Also, it will provide a more discriminative and representative description of objects and scenarios.

4. Our Proposal

GPU Accelerated Shape Description, Extraction and Matching

LISF (Invariant Local Shape Features) [1]



Matches between local shape descriptors in two images. It can be seen how these matches were found even in presence of rotation, scale and translation changes.

• Massively parallel implementation in GPU of the two most time expensive parts of LISF algorithm, i.e. feature extraction and feature matching steps.

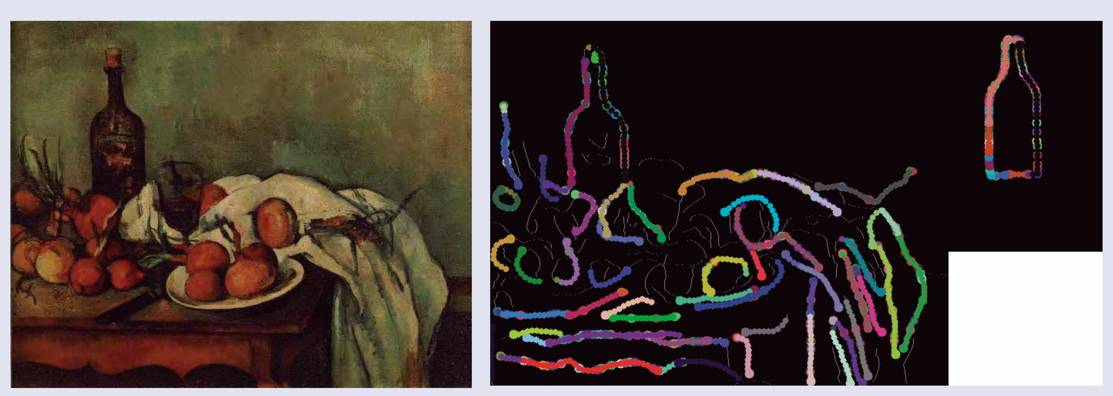
• A 32x speedup is achieved compared to the CPU implementation when extracting features in a 1000 points object contour [2].

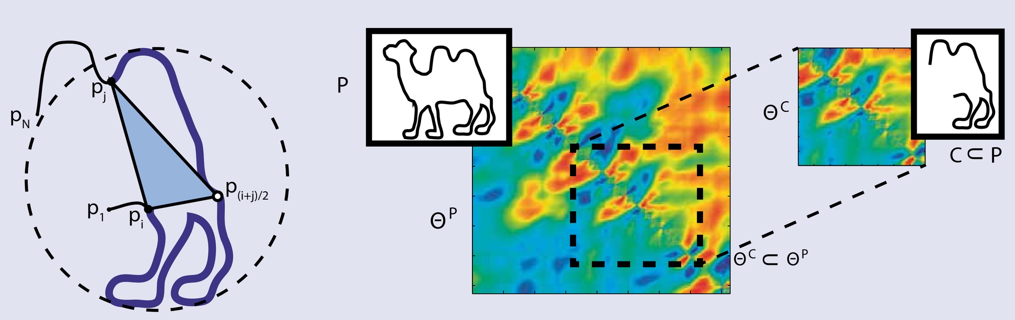
• A 34x speedup is achieved compared to the CPU implementation when matching 290 vs 290 LISF features [3].

• Computed on a NVIDIA GeForce GTX 480.

• Compared to single-threaded Intel CPU Processor at 3.4 GHz.

OCTAR (Open/Closed contours Triangle Area Representation) [4]



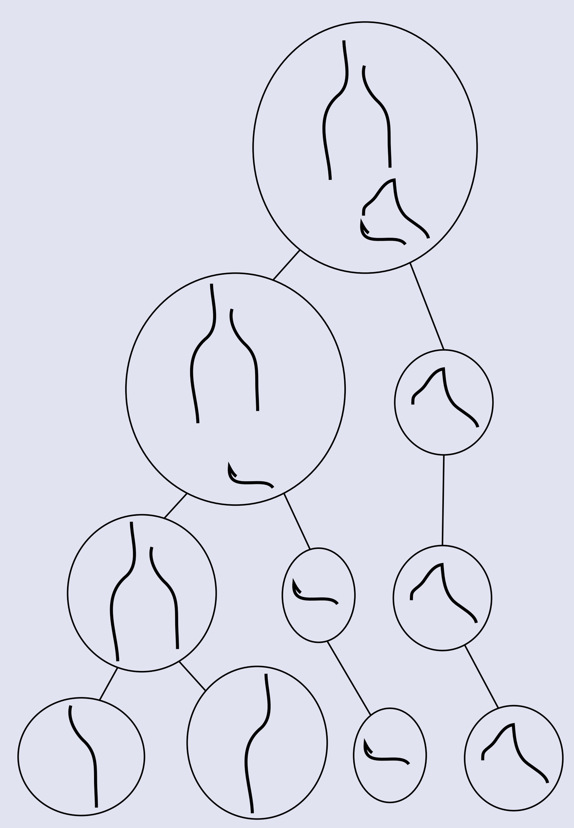


• Open / closed contours.
• Self containing descriptor.
• Invariant to rotation and translation.

• Partial shape matching method robust to partial occlusion and noise in the contour.

Detection Hypotheses

• Combinatorial number of hypotheses are obtained from partial matches.
• Each hypothesis could be evaluated independently.
• We can take advantage of GPU computing in evaluating detection hypotheses in parallel.



Shape-based hypotheses evaluation

• model covering:

$$\mathcal{E}_{COV} = \frac{1}{M}\sum_{i}^{M} w_i$$

• object estimation from hypothesis:

$$O(p) = \begin{cases} \hat{\mathcal{Q}} \rightarrow \hat{\mathcal{F}}, & \text{if } p \in \hat{\mathcal{Q}} \\ p\mathcal{Z}, & \text{if } p \notin \hat{\mathcal{Q}} \end{cases}$$

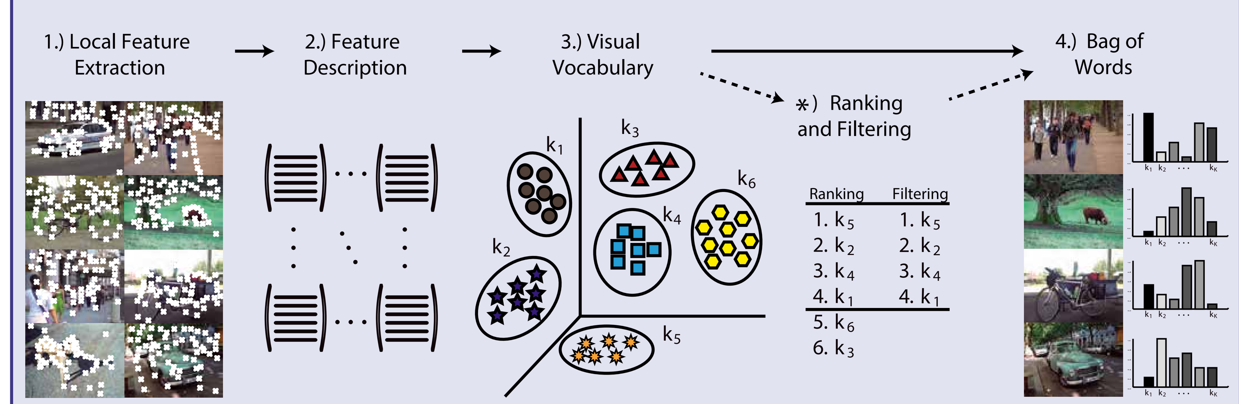
$$\mathcal{E}_{EST} = \Phi(O, \mathcal{Q})$$

• object estimation fit ed to edges:

$$O'(p) = \begin{cases} O(p), & \text{if } p \in \hat{\mathcal{Q}} \\ NN \text{ of } O(p), & \text{if } p \notin \hat{\mathcal{Q}} \end{cases}$$

$$\mathcal{E}_{FIT} = \Phi(O', \mathcal{Q})$$

Appearance-based Hypothesis Evaluation -Improved BoW Representation-



• One of the most widely used appearance-based approaches for representing images for object categorization is the Bag of Words (BoW) approach. Main limitations of the BoW approach:

• The visual vocabulary is built using features that belong to both the object and the background (noise extracted from the background is also considered as part of the object class description).

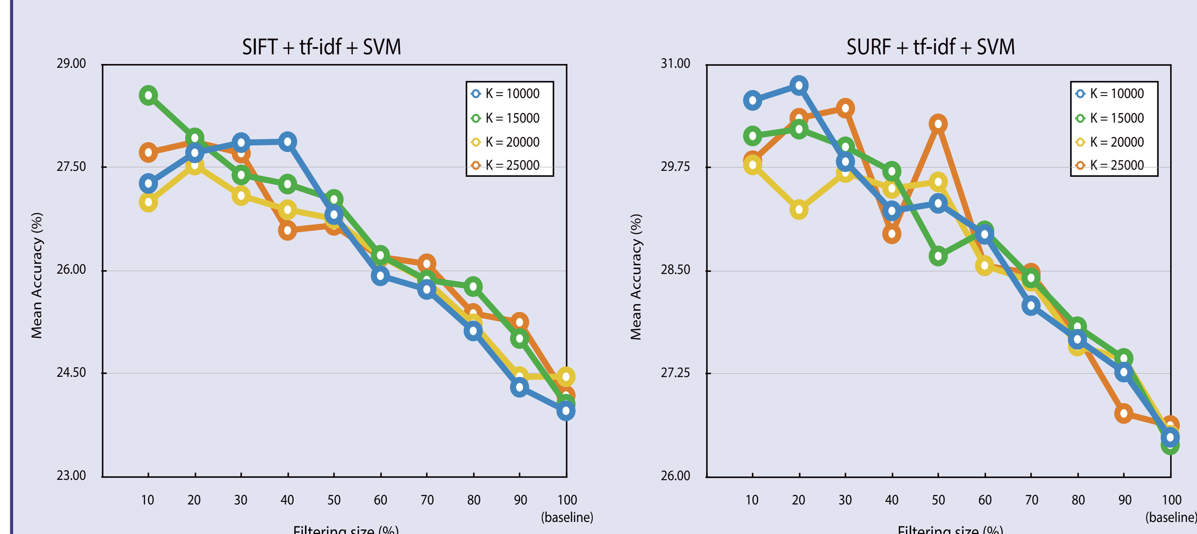
• In the BoW representation, every visual word is used, regardless of its low representativeness or discriminative power.

• There is no consensus about which is the optimal way for building the visual vocabulary.

• We propose three properties to assess the ability of a visual word to represent and discriminate an object class in the context of the BoW approach.

• The visual words that best represent a class, best generalize over intra-class variability and best di° erentiate between object classes will obtain the highest scores for these measures.

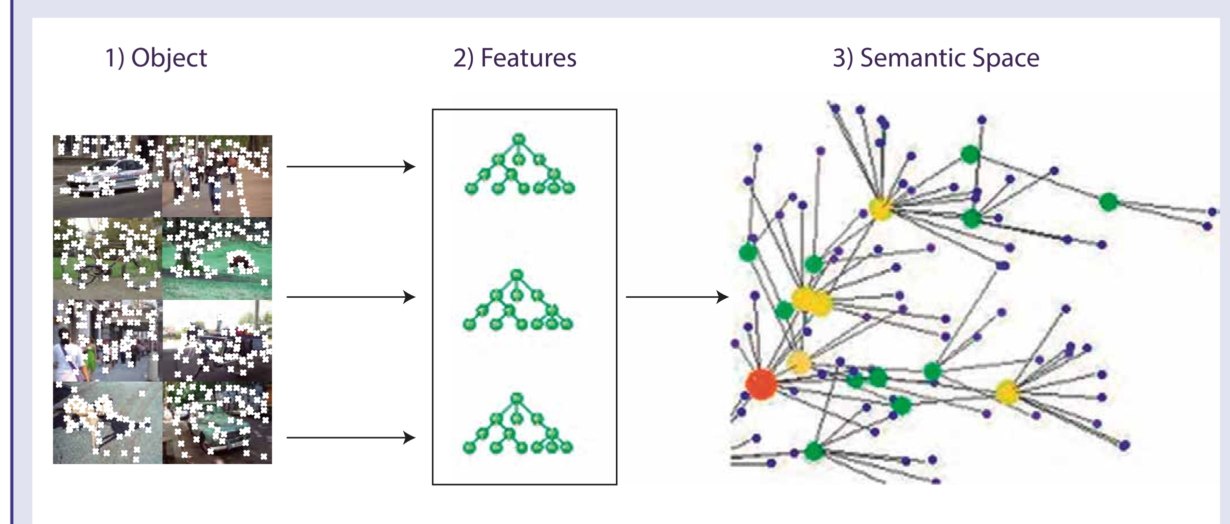
• A methodology for reducing the size of the visual vocabulary based on these properties is also proposed.



• Caltech-101 dataset.
• Vocabularies built with K-means.
• SIFT and SURF local descriptors.
• SVM 10-folds, 10 iterations cross validation.

GPU Accelerated High-Level Semantic Representation

• Our approach computes the BoWs of each feature entry in order to annotate and map semantically based on the similarity features to the same BoWs.

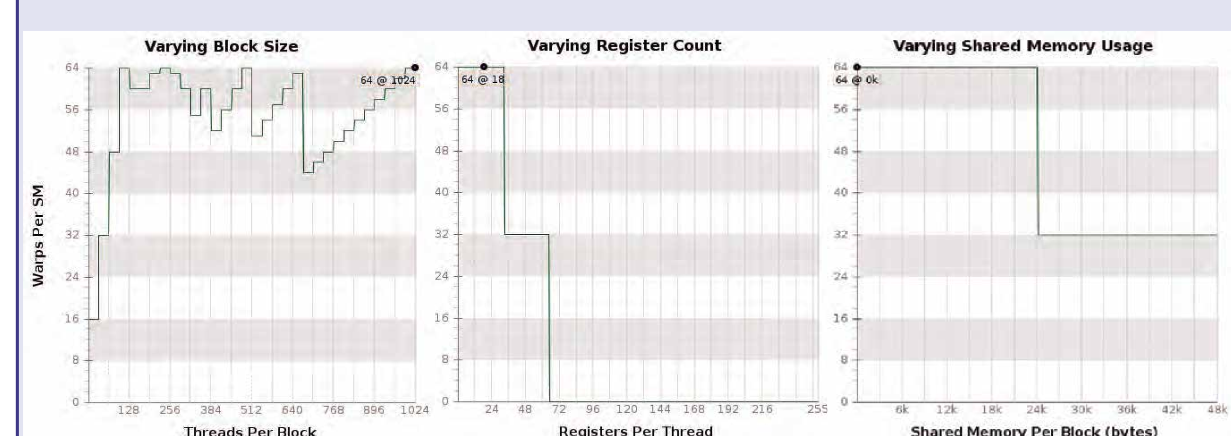


• The metric of success is the distance between object's features and the semantic space that defines it. Semantic analysis can strengthen the object categorization.

• The correlation coefficient measures how correlated the entries are with a value between 1 and -1, where 1 indicates that the entries are perfectly correlated, 0 indicates no correlation, and -1 means they are perfectly inversely correlated.

• Since the number of entries is high, the computation of the features score should be divided into several coprocessors (cores) in order to accelerate the process.

• Each coprocessors of the GPU has as an input: (a) the strength of the feature and (b) the vector with similarity values. Each coprocessor computes the strength of every feature entry until there are no entries left. The score for each feature entry is multiplied by the similarity correlation between all the features.



Varying block size between threads per block and warps per SM. Decrease in the number of registers per thread. Shared memory segment varies depending on the size per block.

References

• [1] L. Chang, M. Arias-Estrada, L. E. Sucar and J. Hernández-Palancar. *LISF: An Invariant Local Shape Features Descriptor Robust to Occlusion*. In Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods, pages 429-437, Angers, France, 2014.

• [2] L. Chang, M. Arias-Estrada, L. E. Sucar, J. Hernández-Palancar. *Efficient Extraction and Matching of LISF Features in GPU*. Poster Session of GTC 2014: GPU Technology Conference 2014.

• [3] L. Chang, M. Arias-Estrada, L. E. Sucar, J. Hernández-Palancar. *Efficient Local Shape Features Matching using CUDA*. Poster Session of GTC 2013: GPU Technology Conference 2013.

• [4] L. Chang, M. Arias-Estrada, J. Hernández-Palancar, L. E. Sucar. *Partial shape matching and retrieval under occlusion and noise*. In Proceedings of 19th Iberoamerican Congress on Pattern Recognition, CIARP 2014.

Acknowledgements

• L. Chang was supported in part by CONACYT-Mexico scholarship No. 240251.

• To NVIDIA for donation cards to University of Guadalajara Cuda Teaching Centre