Object Localization Enhancement by Multiple Segmentations Fusion

Presented by: Ahmed Mounir Gad
Advisor: Ramon Baldrich

UAB
Universitat Autònoma de Barcelona

CVC

Content

- Introduction
  - Segmentation
  - Object localization
- Related work
  - Multiple segmentations
  - Segments neighborhood
- Core approach
  - Generating multiple segmentations
  - Building an improved segmentation
  - Describing and classifying segments
  - Combining regions classifications
  - Refining with a CRF
- Experiments and results
- Conclusions
- Future work

Introduction

Introduction

Image $\rightarrow$ Segmentation $\rightarrow$ Object localization

Image segmentation?

Object localization?

Image segmentation

- Partitioning an image into a set of non-overlapping regions.

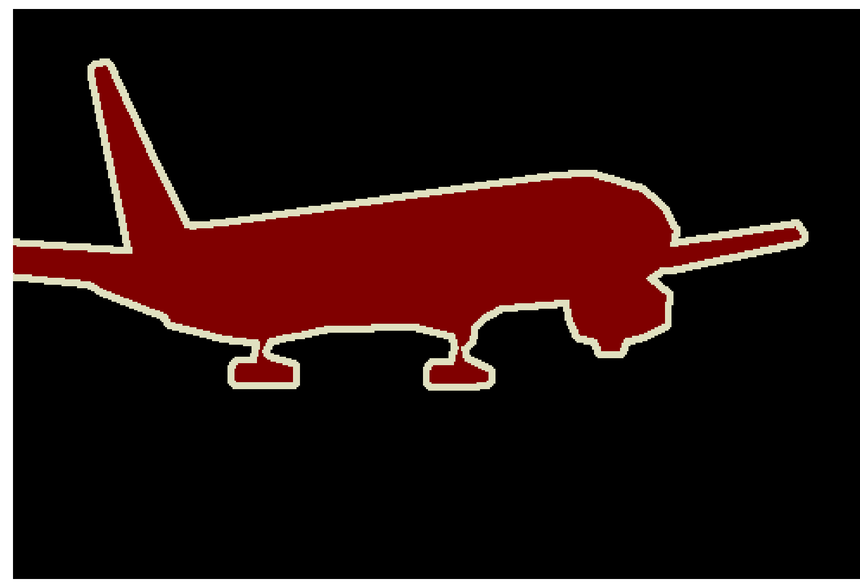
Original Image

Normalized Cut, Seg No = 100

Universitat Autònoma de Barcelona

Object localization

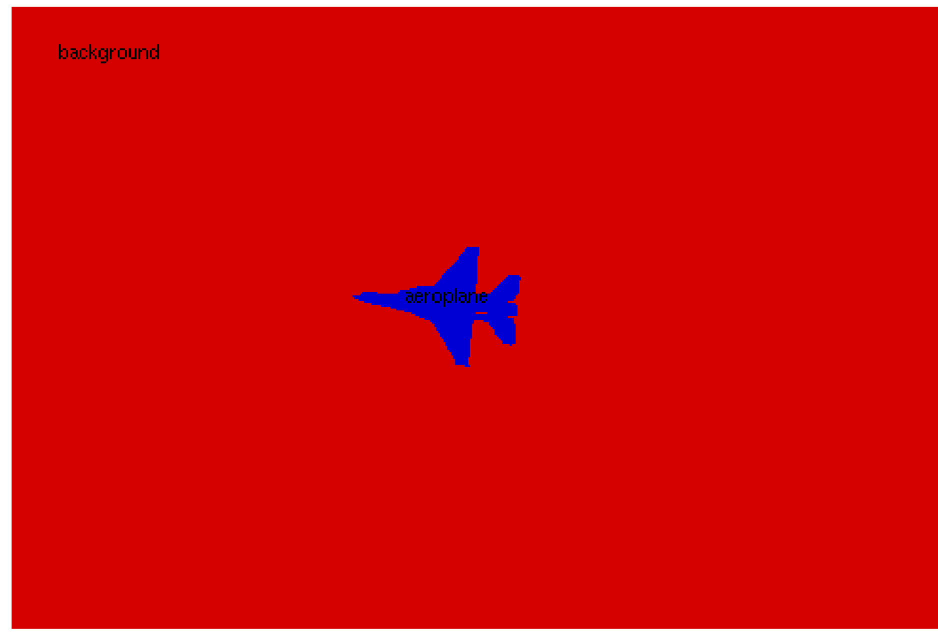
- Recognizing an object inside an image and locating it's exact position



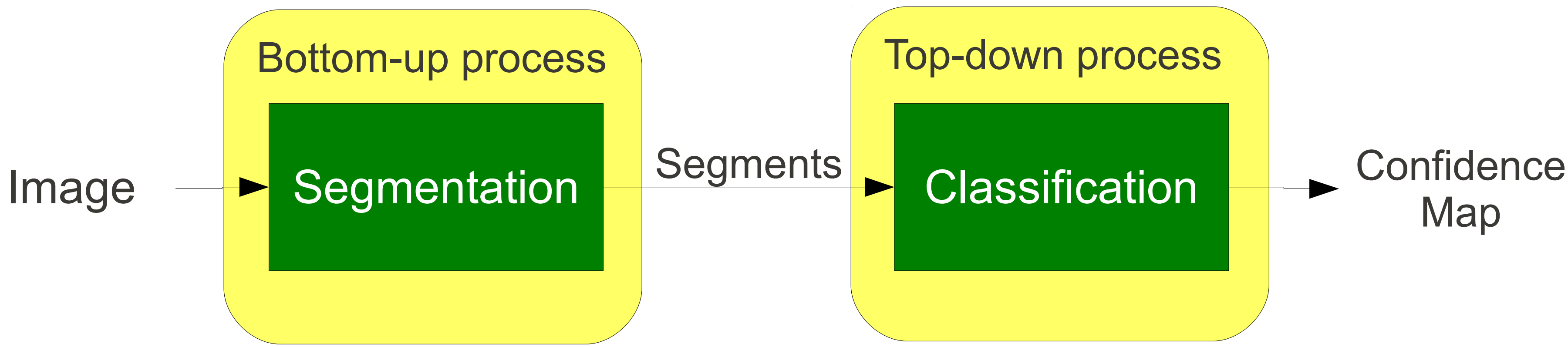
Universitat Autònoma de Barcelona

Object class segmentation

- Producing a pixel-level segmentation of the input image.



Basic framework



Related work

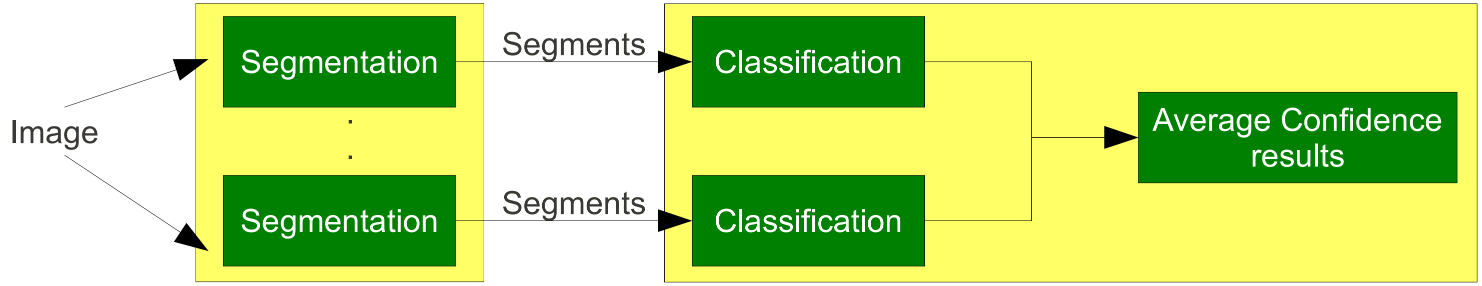
We discuss two techniques that we found useful for our work

- Using multiple segmentations
- Segments neighborhood

Related work – Use of multiple segmentations

Introduced by Pantofaru and Schmid (2008)

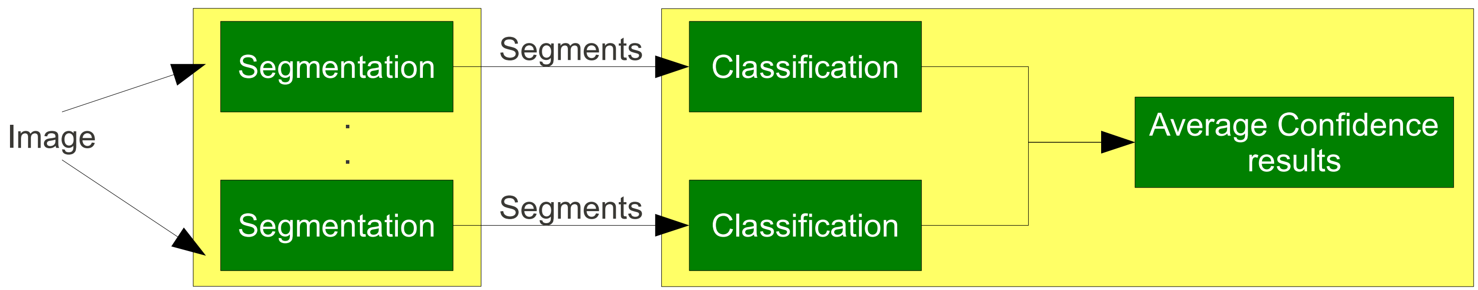
Segments from a single segmentation aren't trustable.



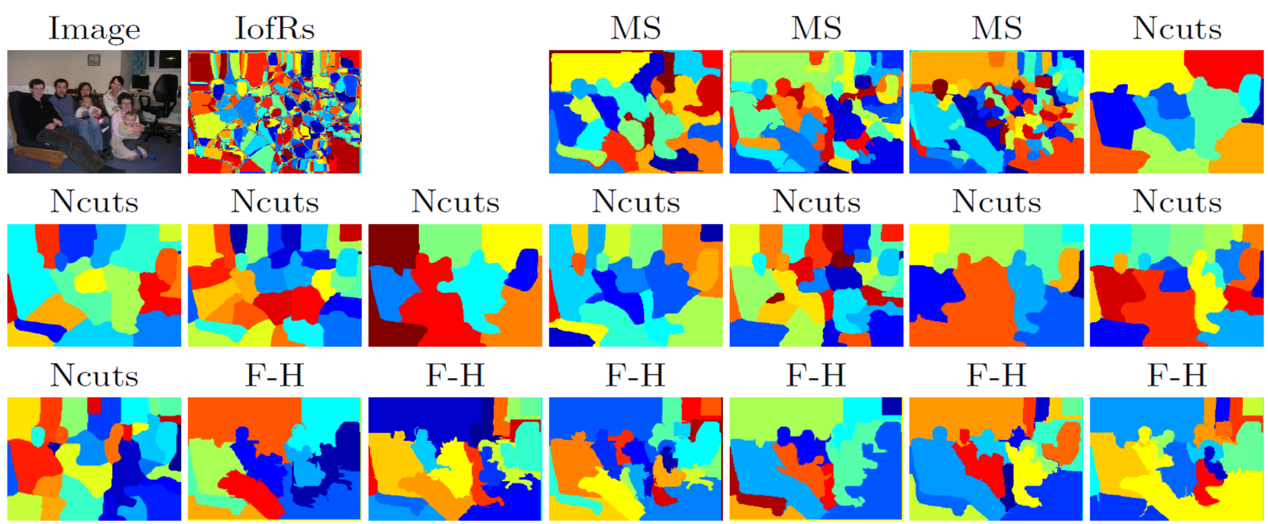
Related work – Use of multiple segmentations

Introduced by Pantofaru and Schmid (2008)

Segments from a single segmentation aren't trustable.



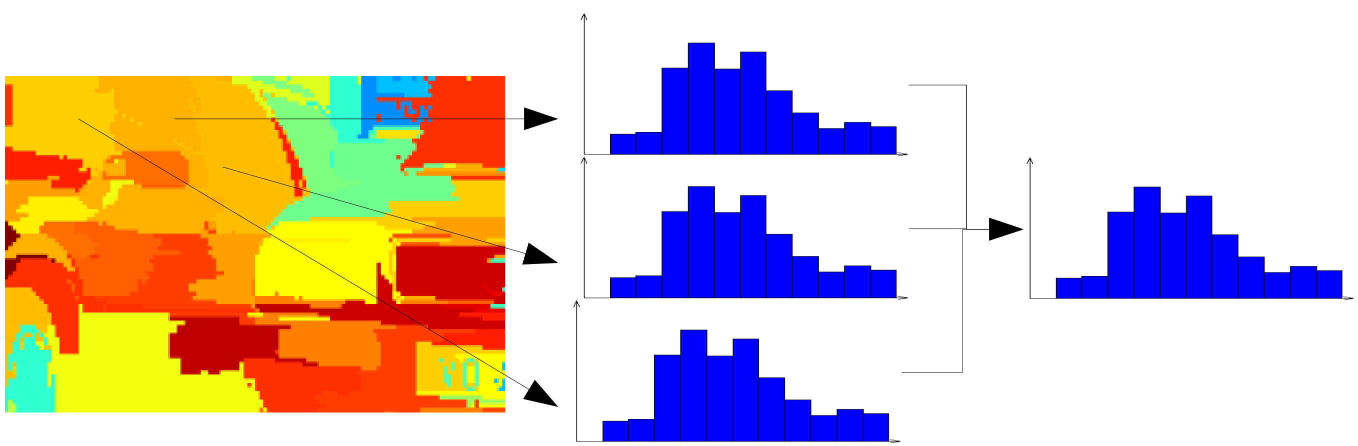
Create Intersection of Regions (IofRs) by intersecting multiple segmentations



Related work – Segments neighborhood

Introduced by Fulkerson et al.(2009)

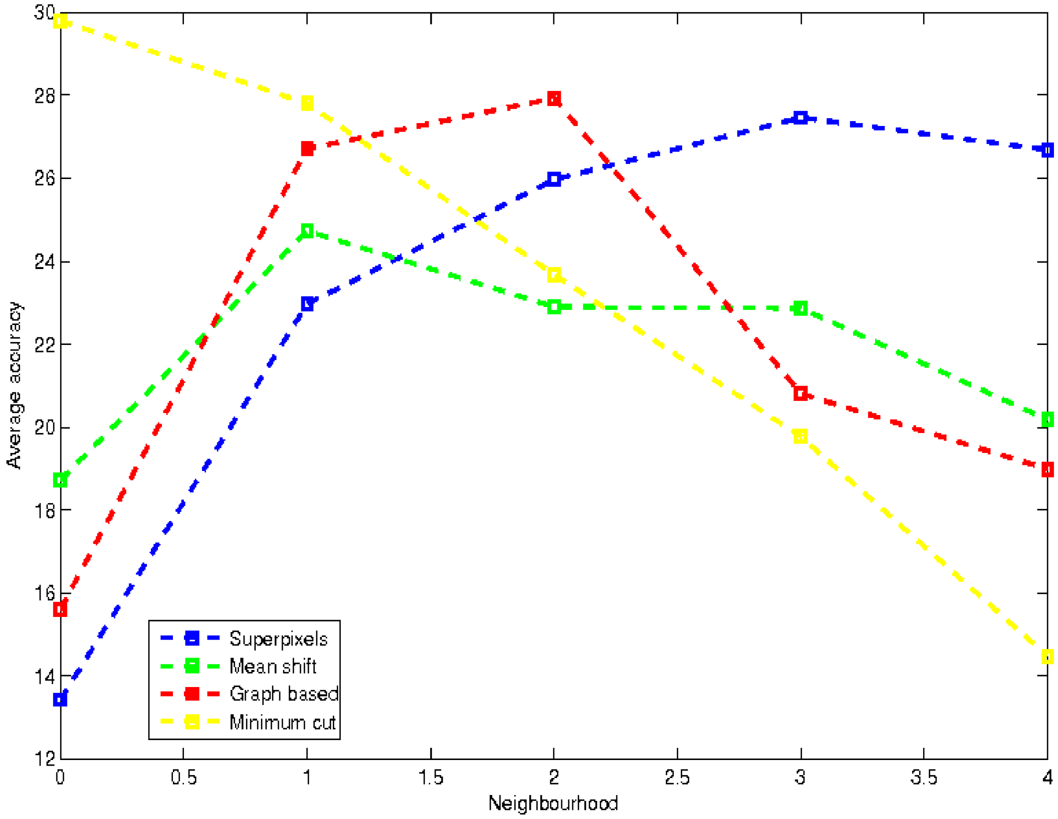
Image → Create Superpixels → Sum neighboring Histograms → Classification → CRF



Related work – Segments neighborhood

Apply on segments instead of superpixels.

On increasing the neighborhood, it means considering a larger segment.



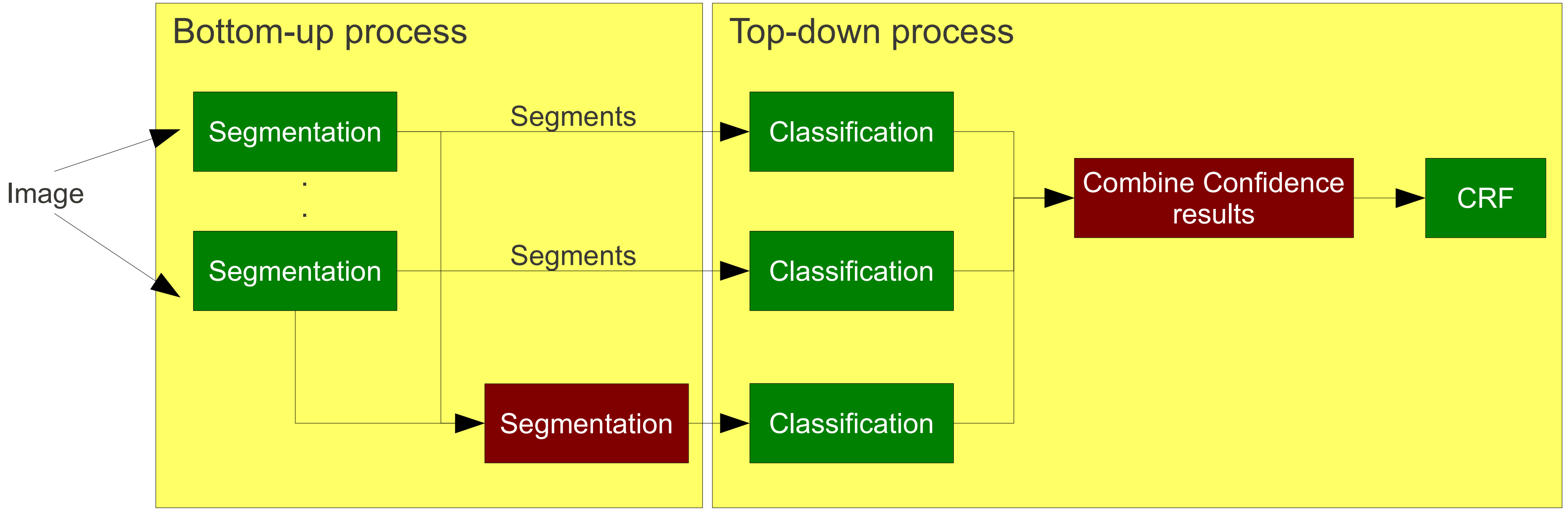
Average segments count:

Graph-Based:	2444
Superpixels:	1201
Mean-shift:	393
Normalized-cut:	100

Observations

- Superpixels aren't well descriptive for the objects (Case of absent neighborhoods).
- At a certain limit, the neighborhood operator fails to give a better performance and results start decreasing.

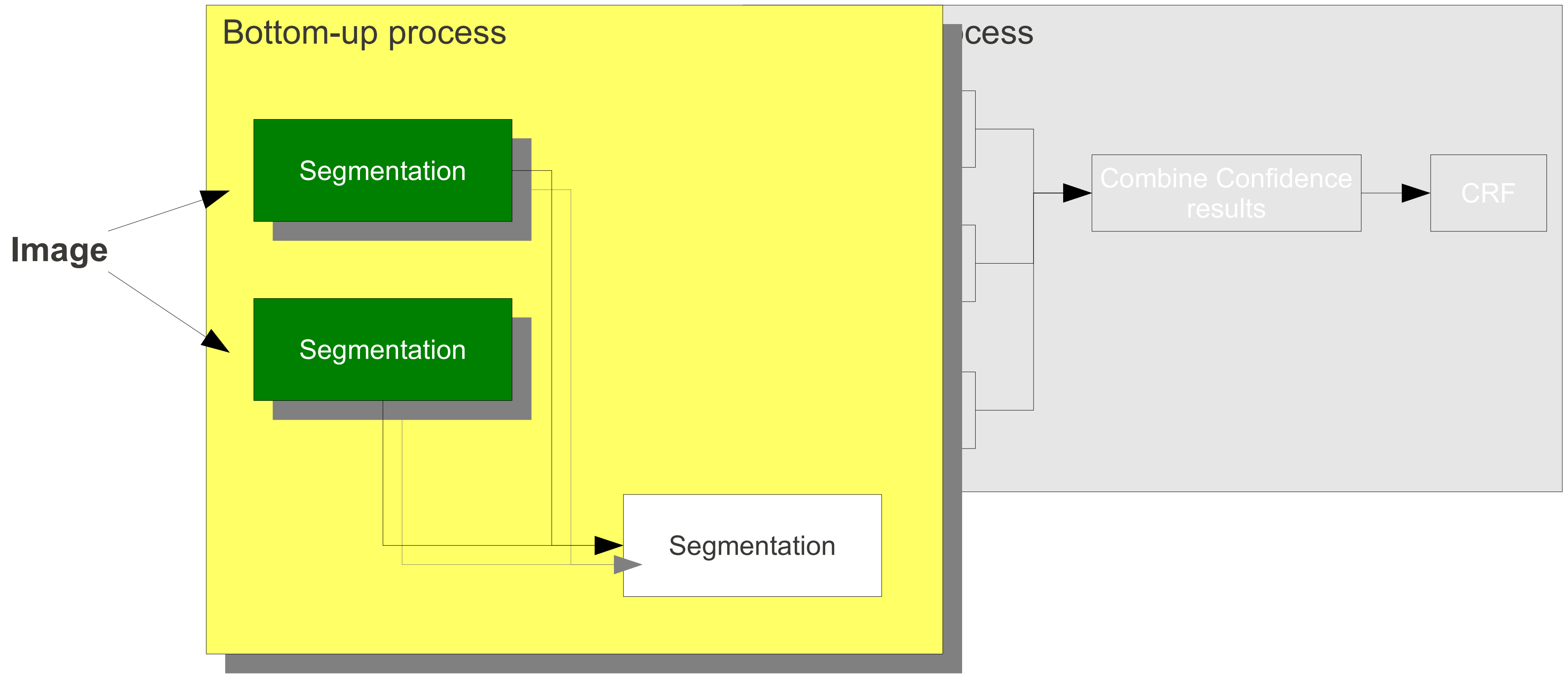
Our framework



Our framework is based on the VLBLOCKS by Fulkerson et al. (2009)

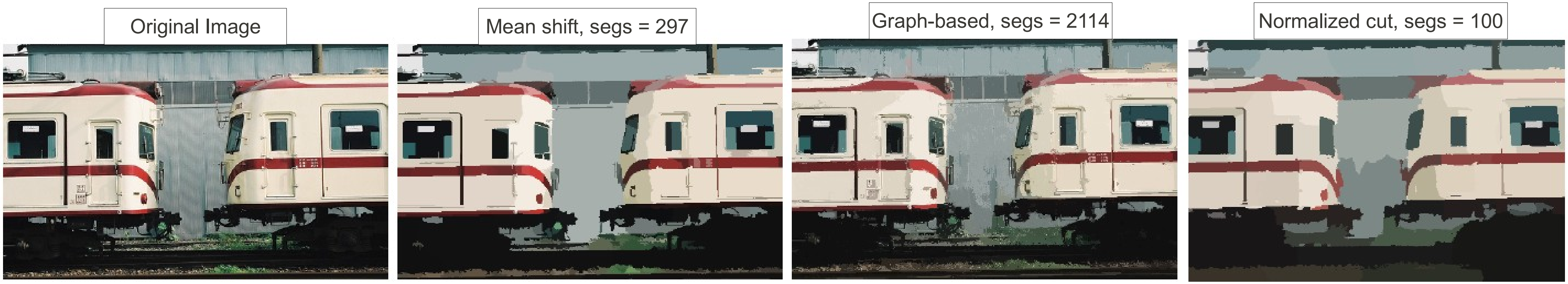
Universitat Autònoma de Barcelona

Our framework

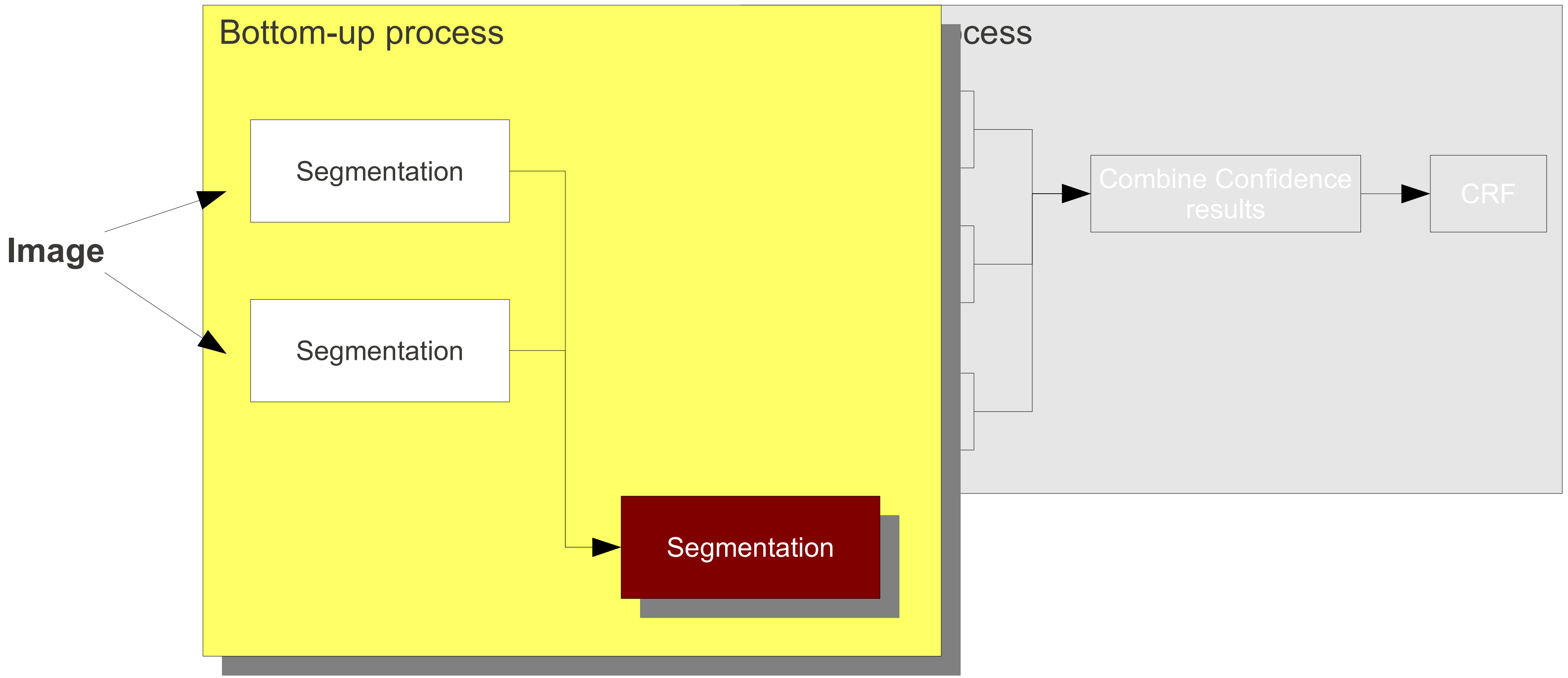


Generating multiple segmentations

Generate multiple segmentations that use different cues.



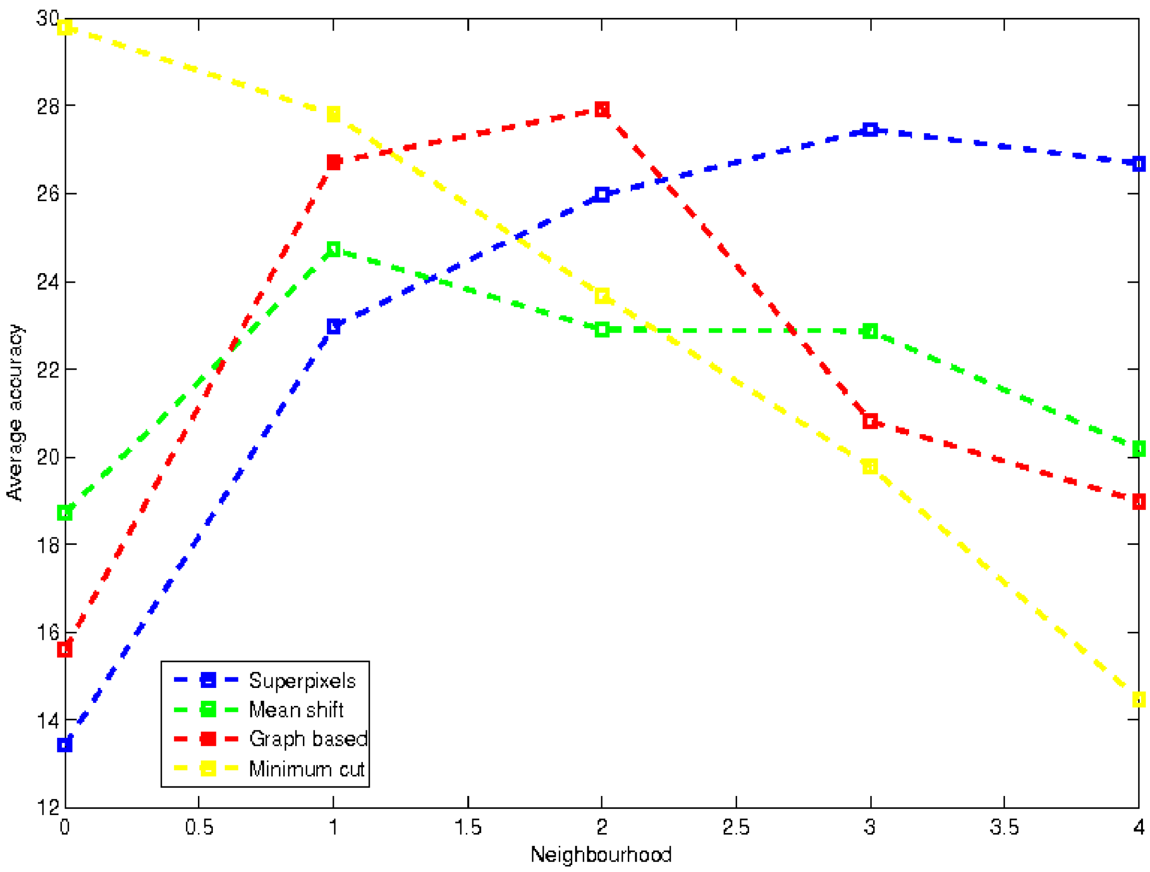
Our framework



Universitat Autònoma de Barcelona

Combining segmentations

Recall this figure



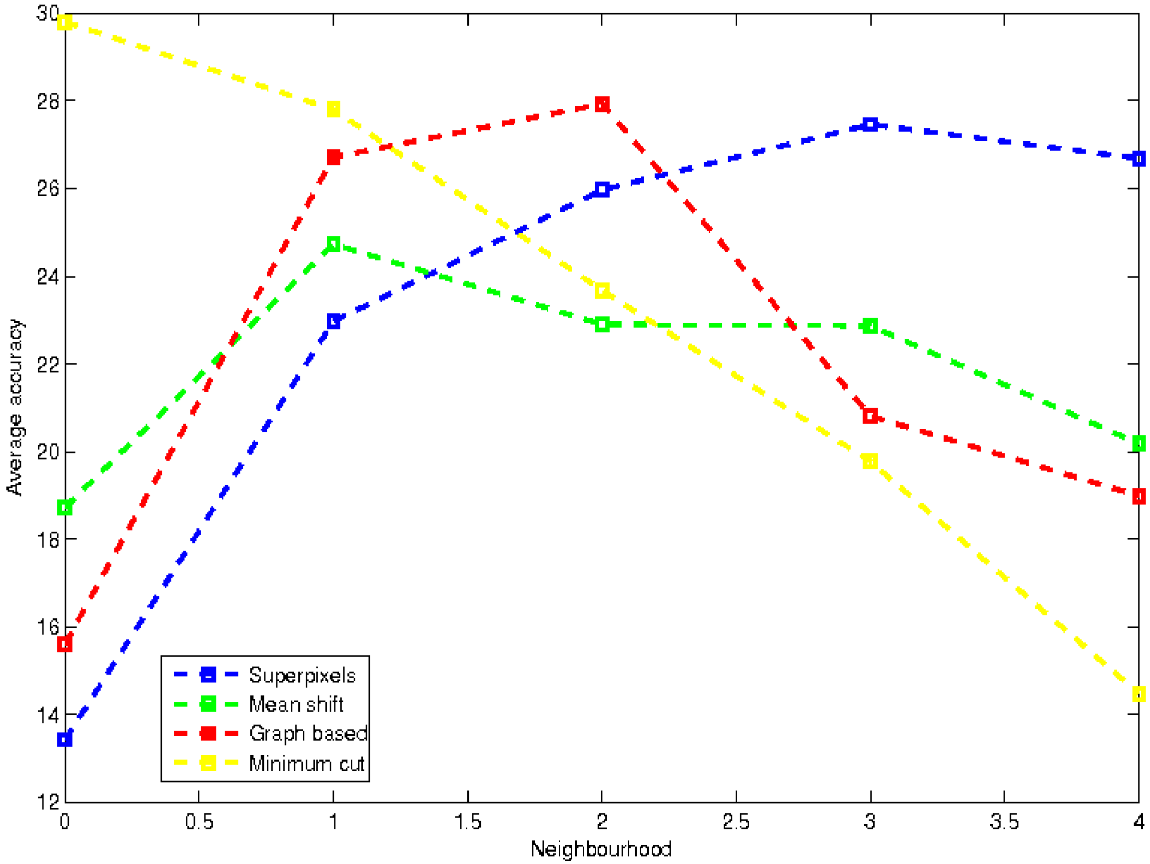
Average segments count:

Graph-Based:	2444
Superpixels:	1201
Mean-shift:	393
Normalized-cut:	100

Universitat Autònoma de Barcelona

Combining segmentations

Recall this figure



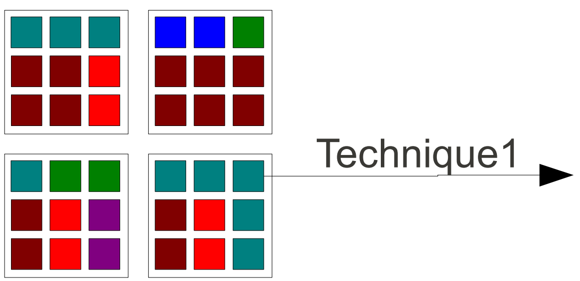
Average segments count:

Graph-Based:	2444
Superpixels:	1201
Mean-shift:	393
Normalized-cut:	100

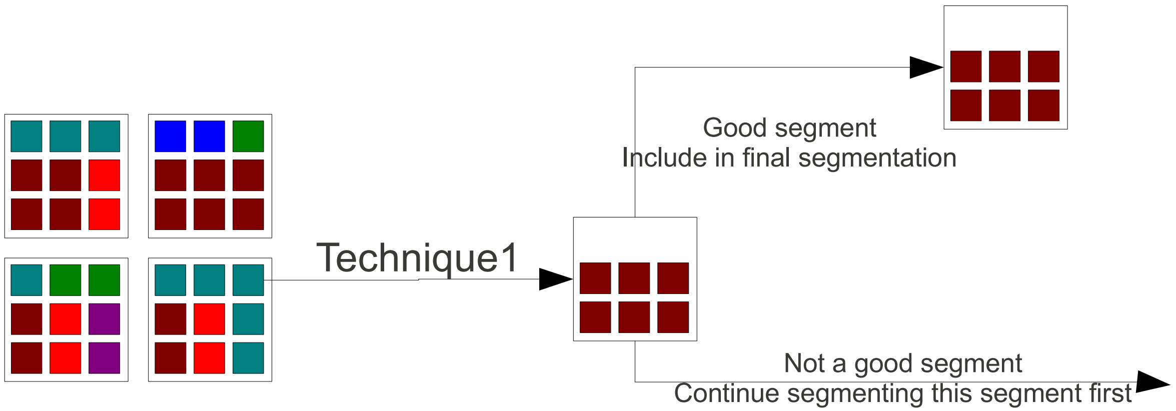
Getting higher accuracy:

- Large segments
- Good segments

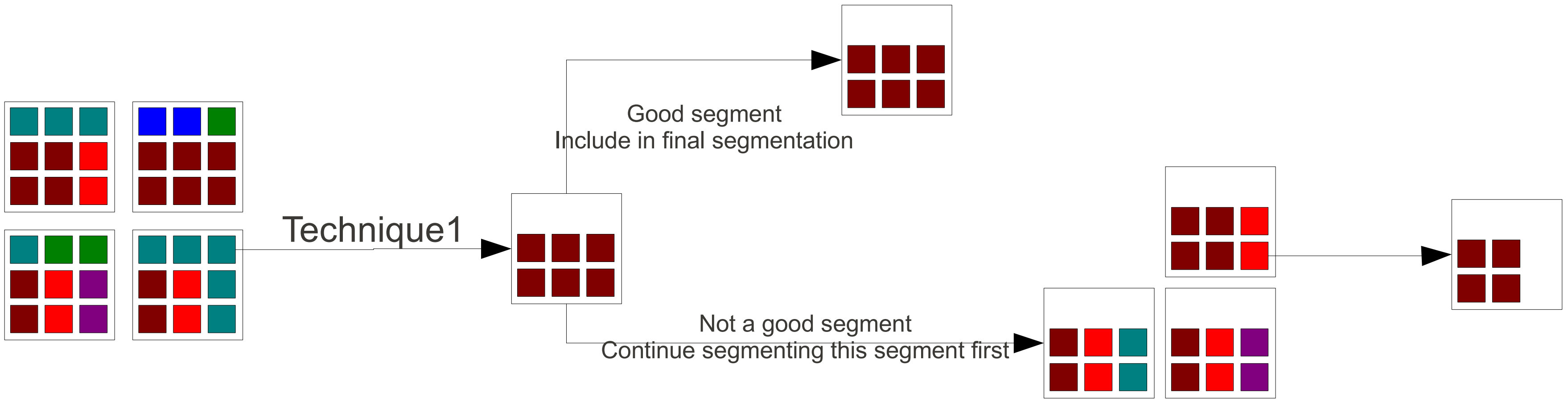
Combining segmentations



Combining segmentations



Combining segmentations



Good segments evaluation

Is this a good segment?

What do we mean by a good segment?

Good segments evaluation

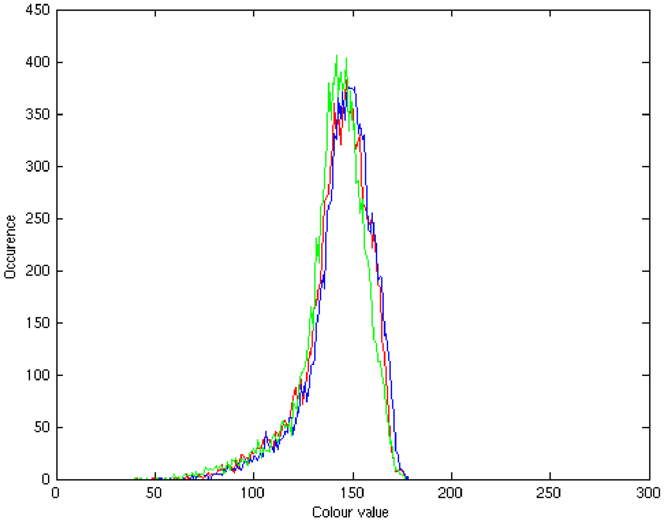
Is this a good segment?

What do we mean by a good segment?

- The largest set of connected pixels that lie in the same class and are coherent in color distribution.

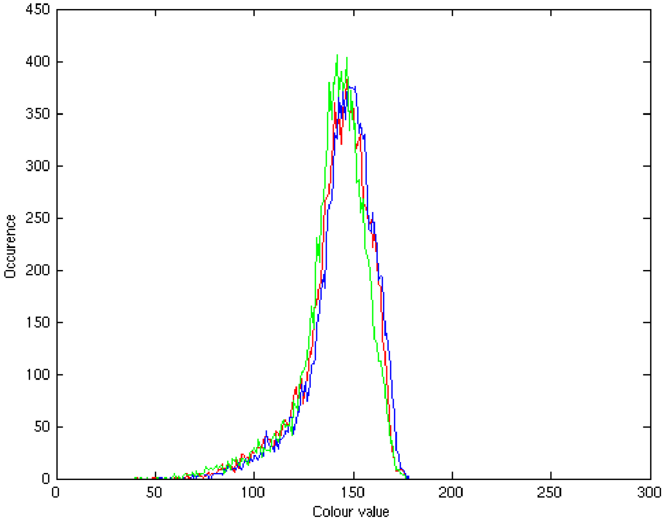
Good segments evaluation

Is this a good segment?



Good segments evaluation

Is this a good segment?



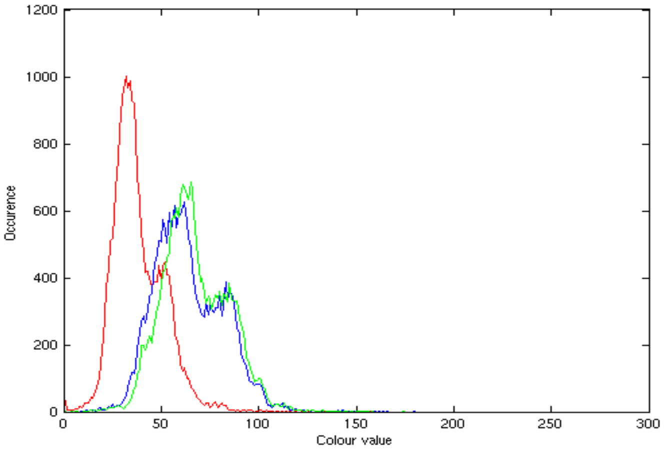
Normality test (Jarque-Bera test)

Unimodality test (Dip test)

Good segments evaluation

Is this a good segment?



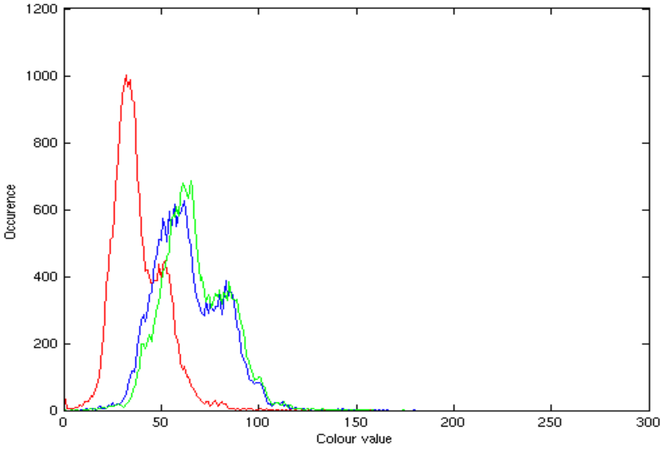


Normality test (Jarque-Bera test)

Unimodality test (Dip test)

Good segments evaluation

Is this a good segment?



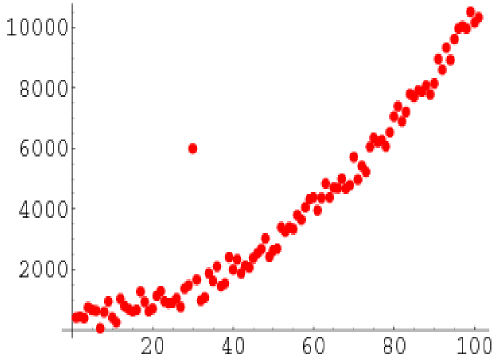
Normality test (Jarque-Bera test)

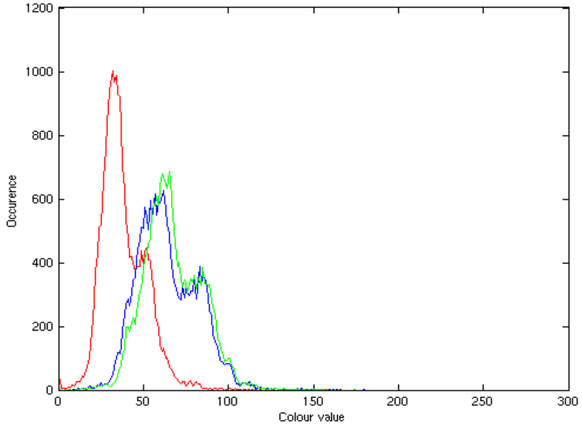
Unimodality test (Dip test)

Outliers detection, RAD, alternative segmentation

Evaluation using outliers detection

Outlier: Observation that deviates markedly from other members of the sample.





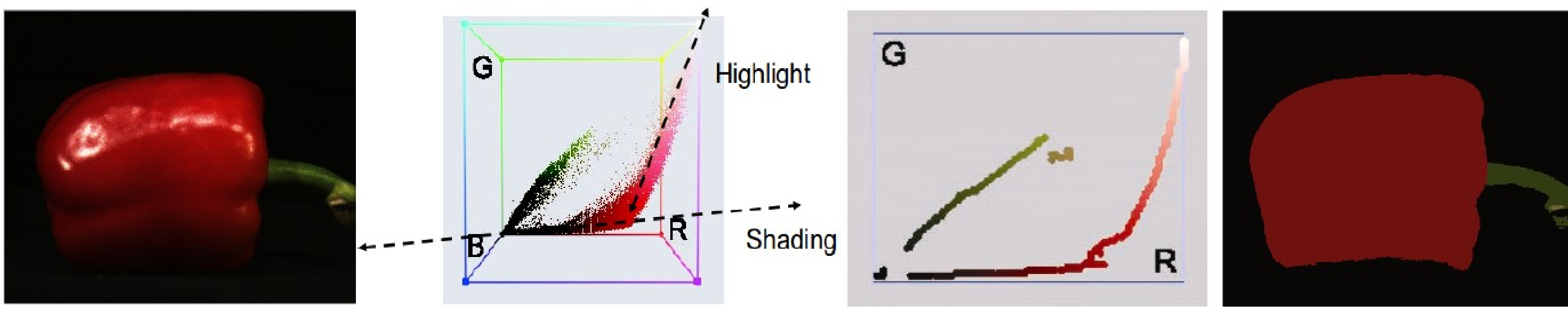
We used the Z-test to detect outliers

Problem: Shadows and illumination

Find another method robust to shadows and illumination

Evaluation using RAD

RAD segmentation introduced by Vazquez et al. (2008)



$$f(x) = m^b(x)c^b + m^i(x)c^i$$

A good segment is a segment whose histogram follows only one ridge

Evaluation using alternate segmentation

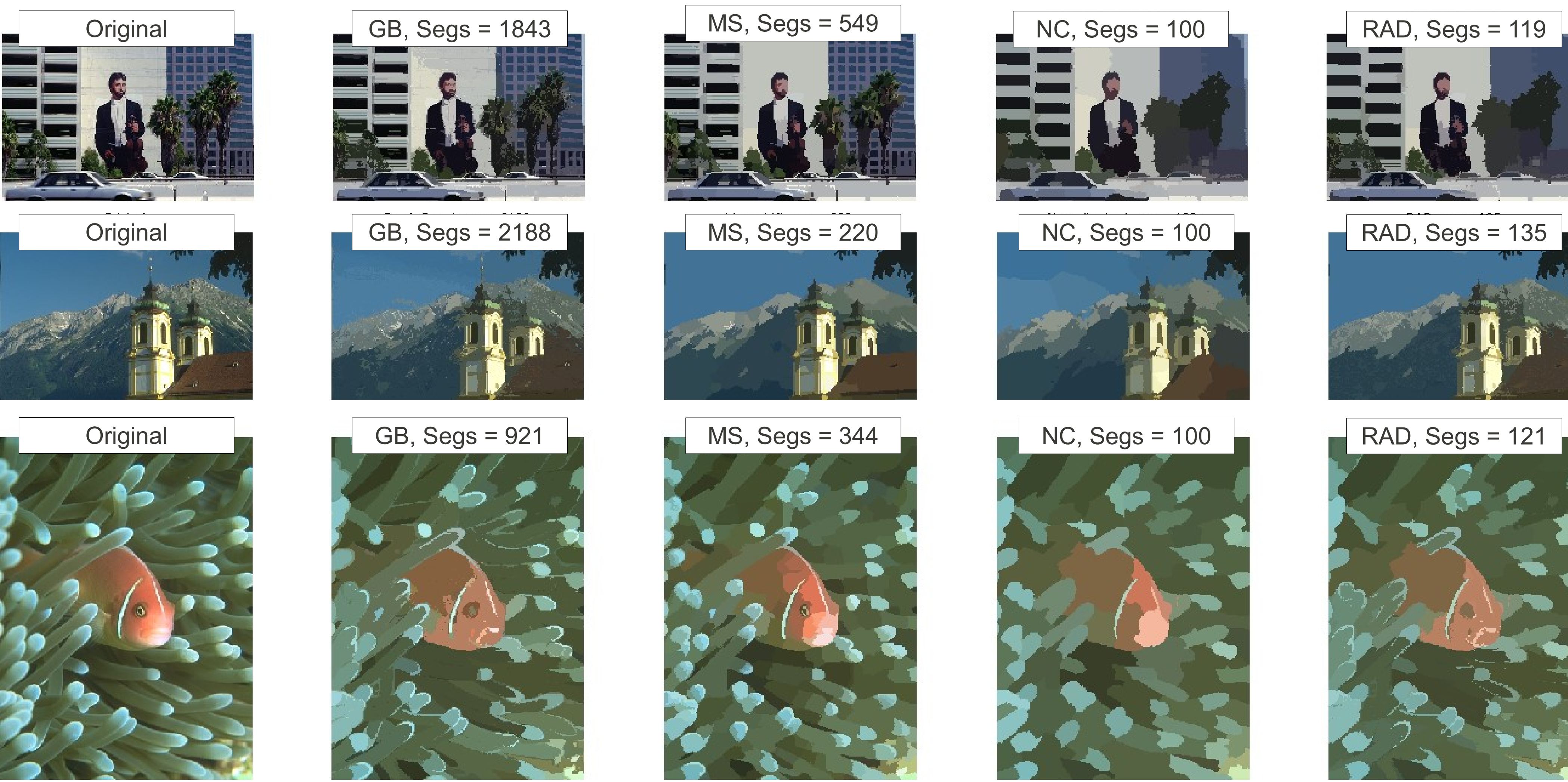
Use alternate segmentations that use different cues to verify segment goodness.

Example: To make sure a segment from Mean-Shift is a good segment, segment it using Graph-Based techniques.

Examples of results

Universitat Autònoma de Barcelona

More results (Berkeley Dataset)



Universitat Autònoma de Barcelona

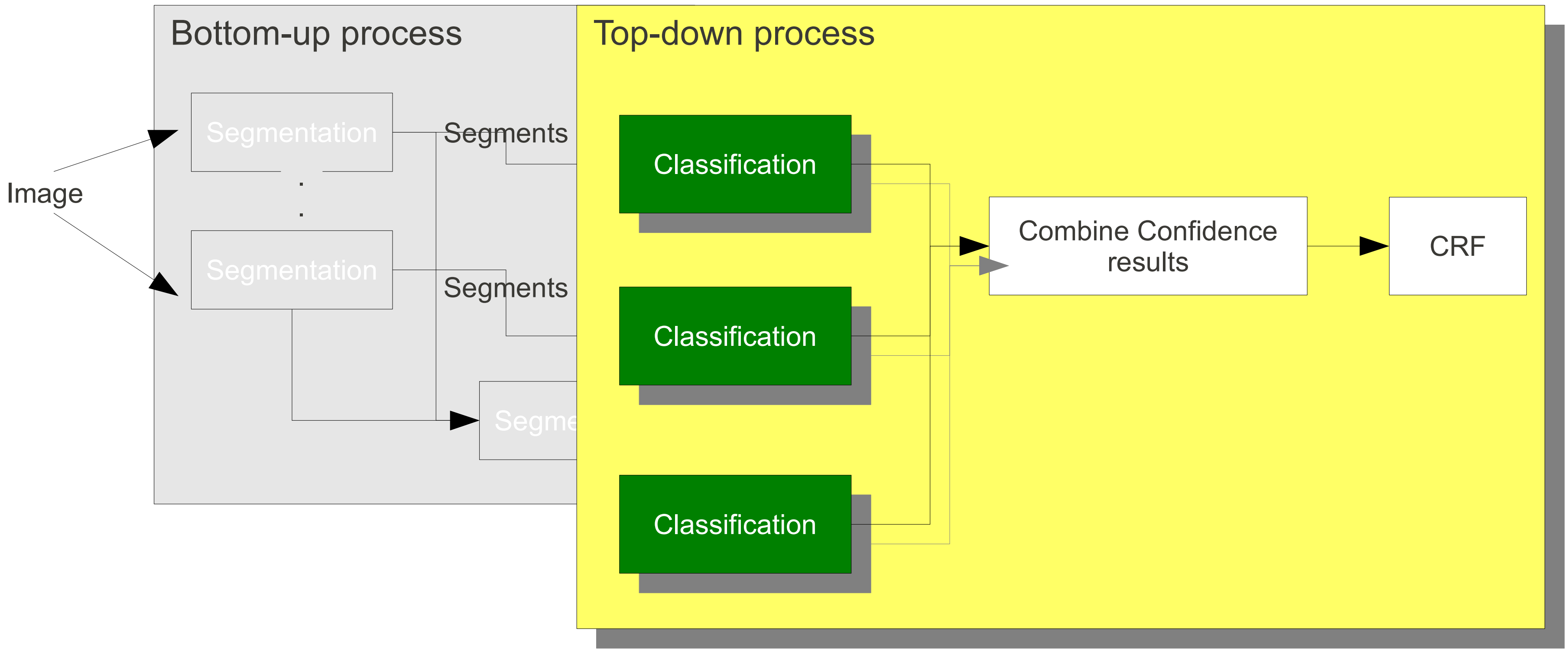
Berkeley Segmentation Dataset and Benchmark

100 images, compared to 5 human segmentations

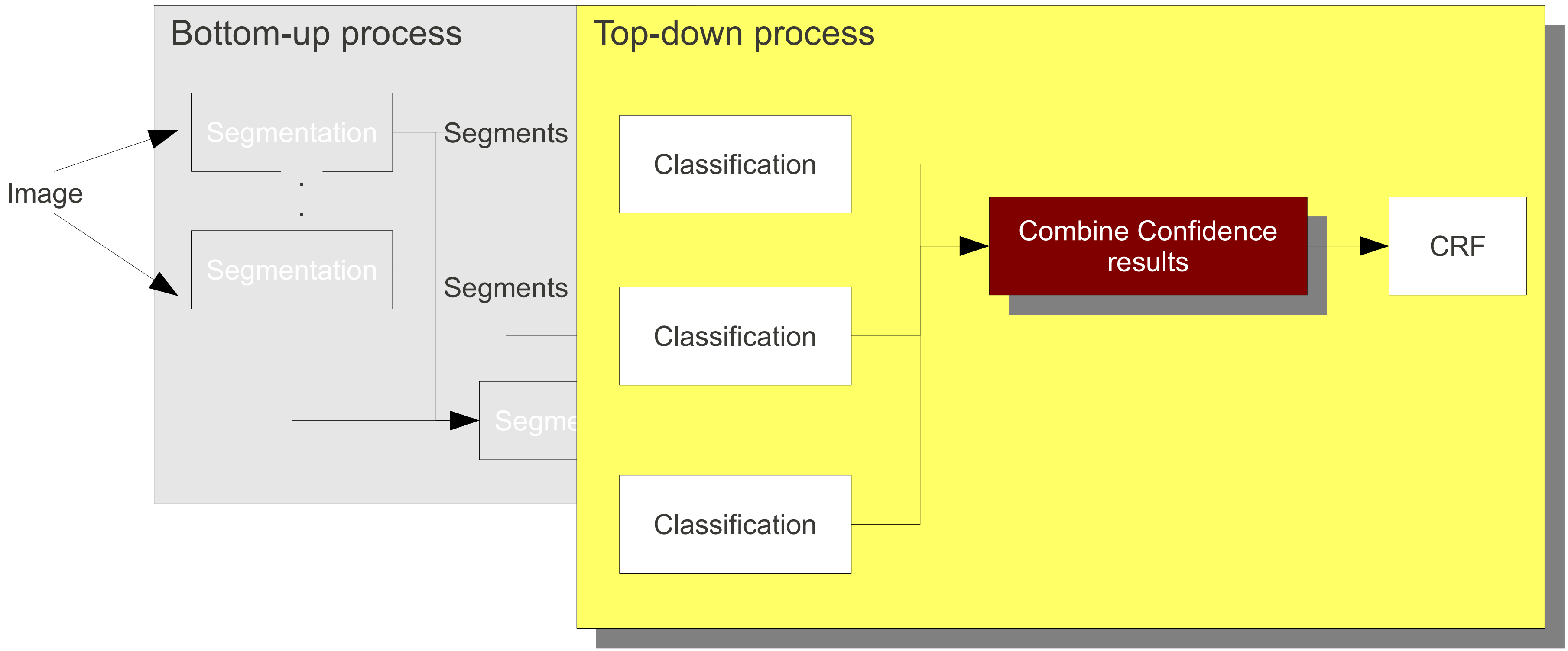
Better results in 2 out of 4 benchmarks

	PRI	VoI	GCE	BDE
Mean-Shift	0.7424	4.5293	**0.0842**	**14.2716**
Graph-Based	0.7082	5.1148	0.1150	17.2428
Normalized-cut	0.7079	4.1370	0.1153	14.7337
Mix+RAD	**0.7483**	**3.5364**	0.1433	14.8047

Our framework



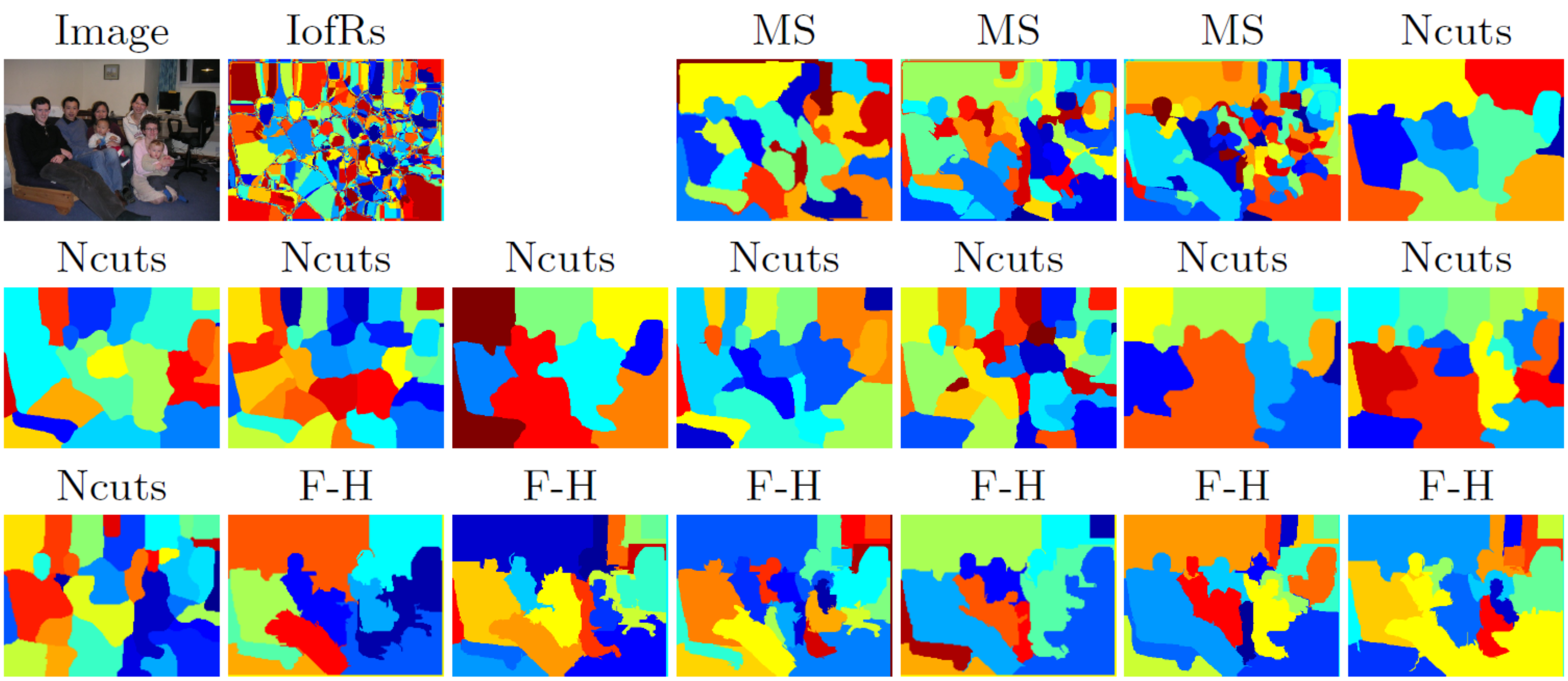
Our framework



Universitat Autònoma de Barcelona

Combining regions classifications

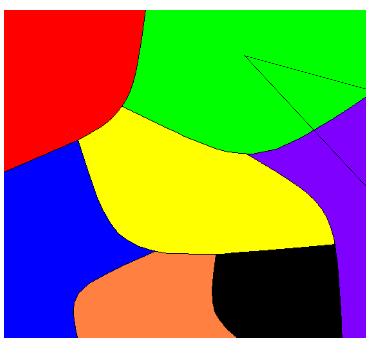
Recall the IofR concept by Pantufaro et al. (2008)



Instead of averaging the confidences, we use our proposed voting technique.

Voting technique

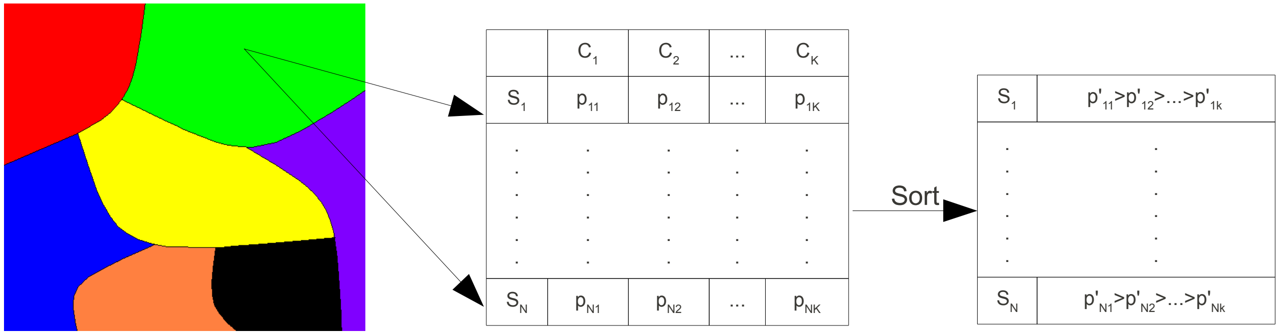
For every IofR.



	C_1	C_2	...	C_K
S_1	p_{11}	p_{12}	...	p_{1K}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S_N	p_{N1}	p_{N2}	...	p_{NK}

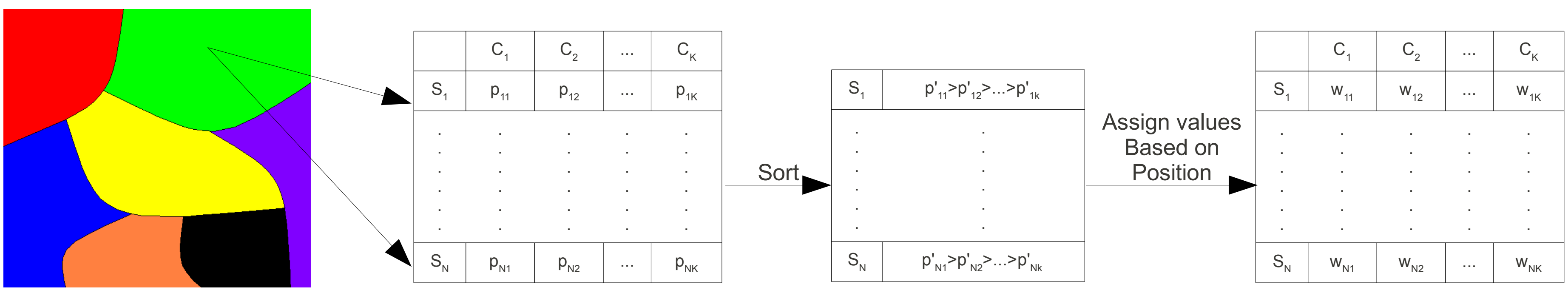
Voting technique

For every IofR.



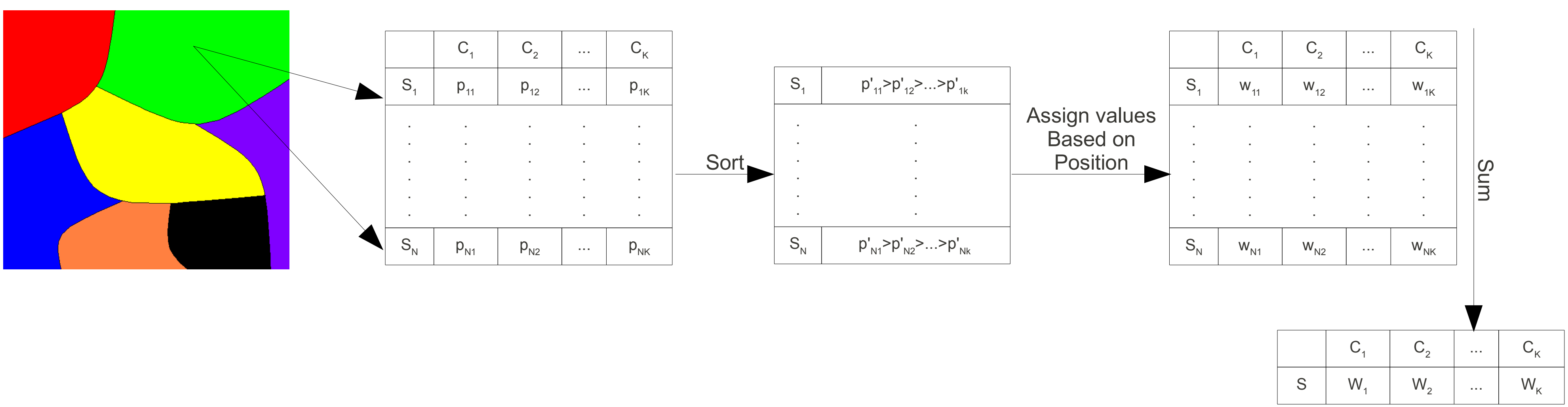
Voting technique

For every IofR.



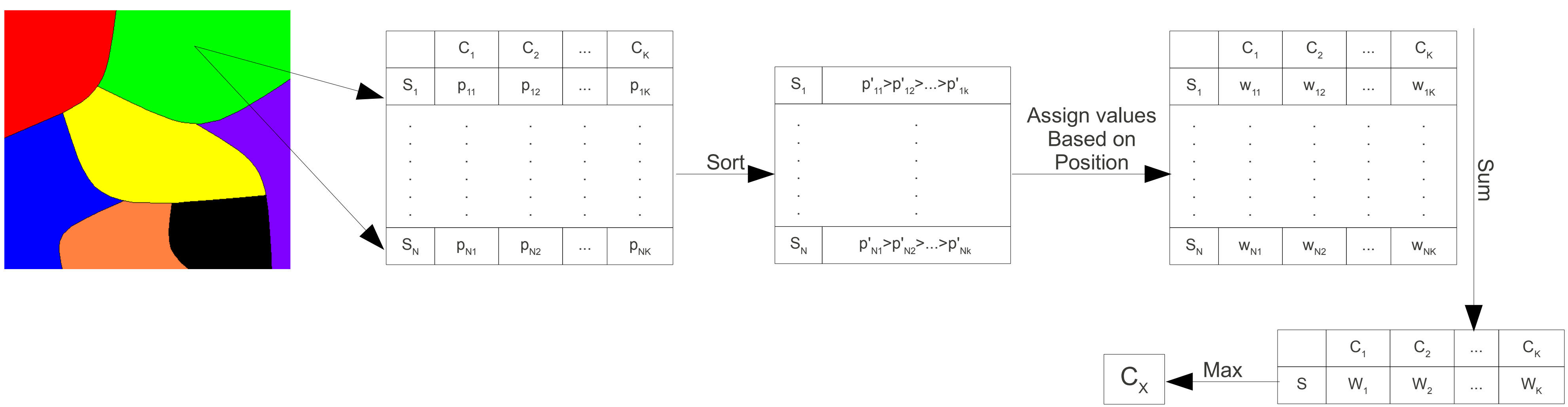
Voting technique

For every IofR.

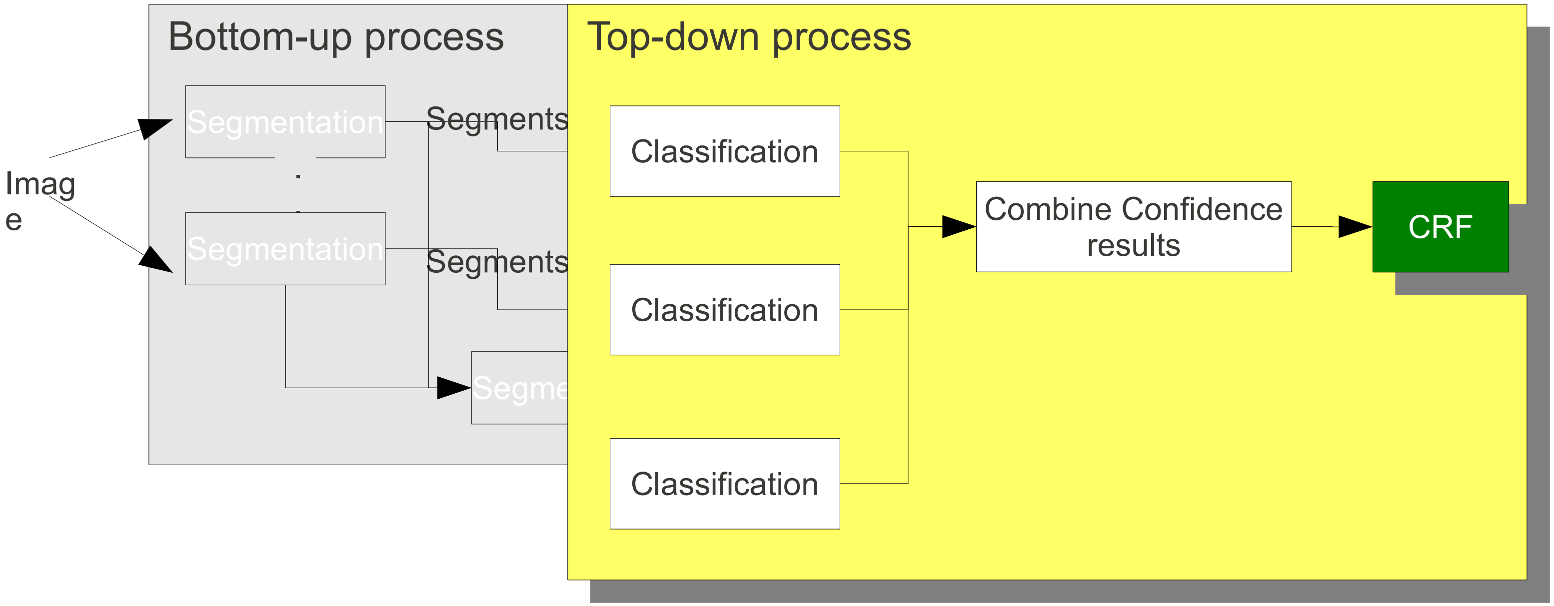


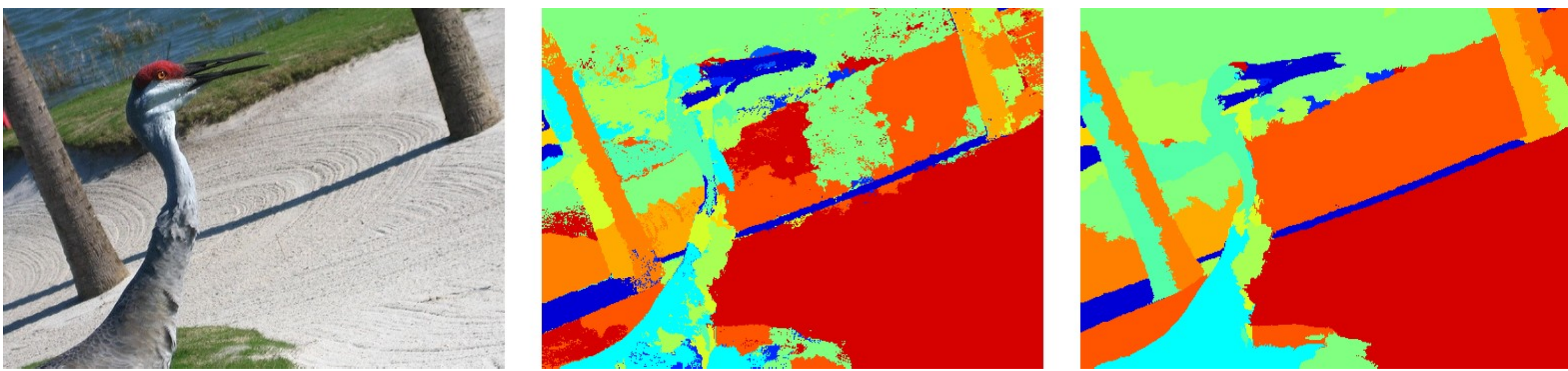
Voting technique

For every IofR.



Our framework





Experiments and results

Experimental setup

PASCAL VOC 2007 segmentations dataset.

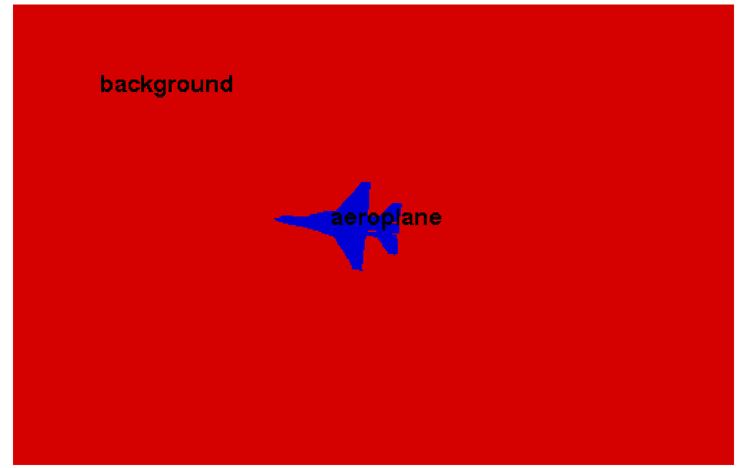
632 images. 422 for training, 210 for testing.

Class accuracy:

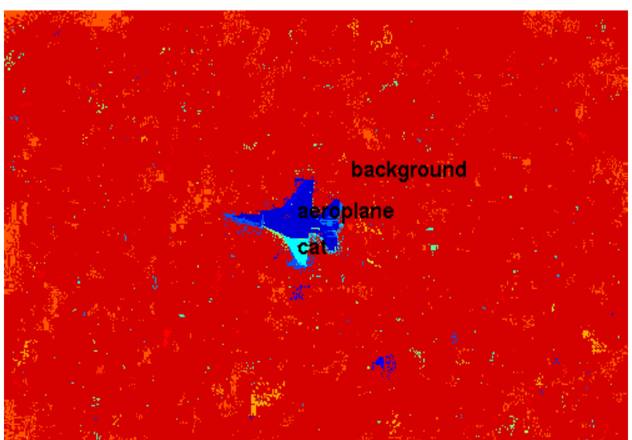
$$\frac{correctlyclassifiedpixels}{groundtruthpixels+misclassifiedpixels}$$

Qualitative results

Votes CRF

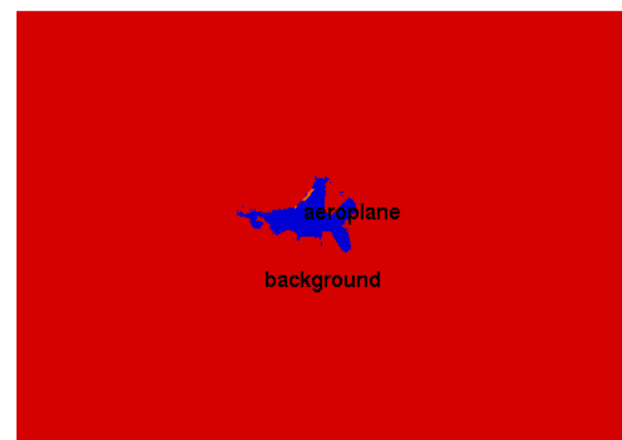


Graph based



Mean shift

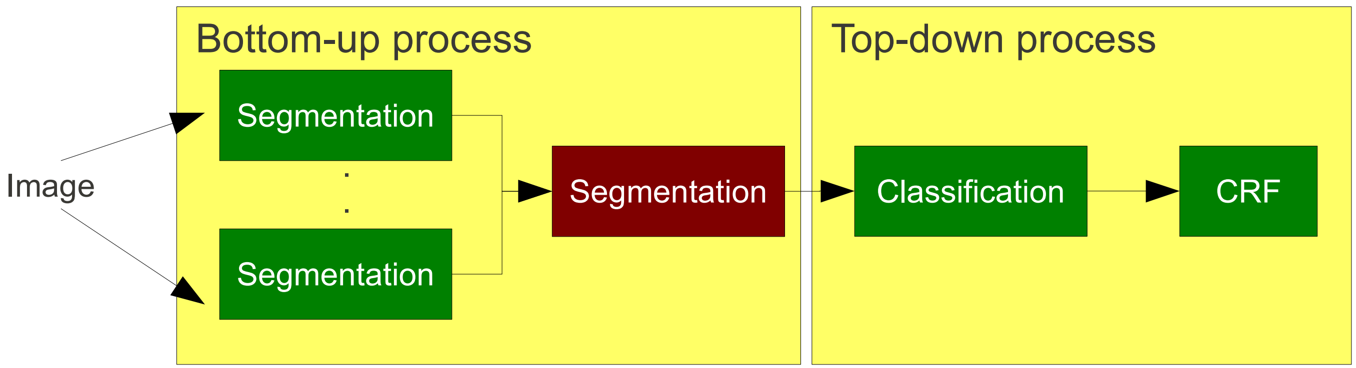
Normalized cut



Superpixels versus meaningful segments

	Background	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Dinningtable	Dog	Horse	Motorbike	Person	Pottedplant	Sheep	Sofa	Train	TVmonitor	Avg Accuracy	Avg segcount
Superpixels	14	21	7	7	22	12	6	7	26	**12**	12	9	10	10	11	2	7	18	35	13	22	13	1201
Mean-shift	16	**25**	12	8	16	7	15	15	**52**	10	9	12	**20**	5	19	16	23	**23**	**43**	17	31	19	393
Graph-based	18	21	20	2	17	6	11	21	39	2	15	14	3	5	37	8	**33**	5	7	29	15	16	2444
Normalized cut	**37**	17	**25**	**24**	**27**	**16**	**41**	**38**	50	9	**26**	**42**	12	**15**	**66**	**30**	10	19	27	**58**	**35**	**30**	**100**

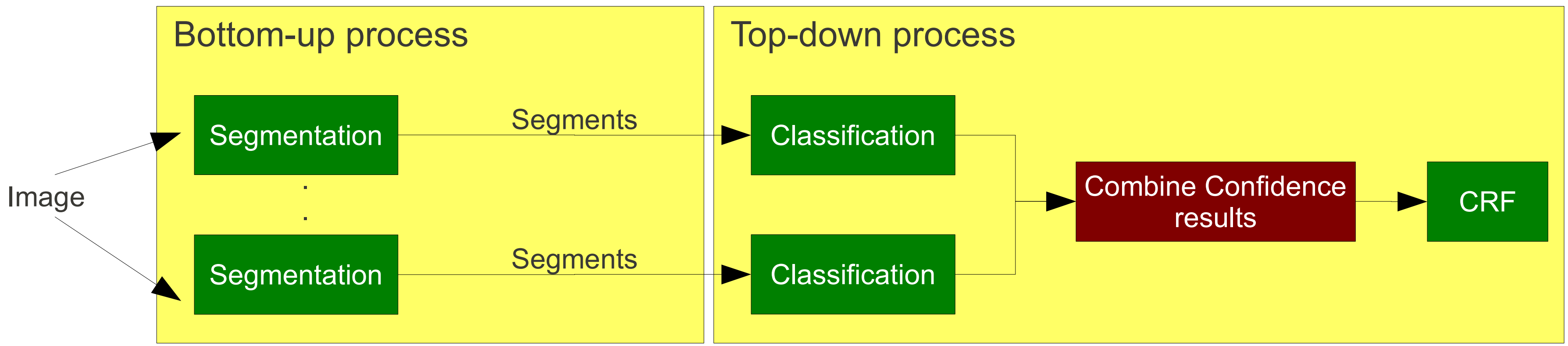
Our new segmentation method



	Background	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Dinningtable	Dog	Horse	Motorbike	Person	Pottedplant	Sheep	Sofa	Train	TVmonitor	Avg Accuracy	Avg segcount
Outliers	22	9	15	8	**25**	14	18	11	43	11	13	16	**31**	**22**	27	6	31	15	**35**	27	18	20	411
AltSeg	10	**33**	17	3	2	**22**	14	9	44	7	10	31	5	7	20	2	24	10	8	7	20	15	1769
RAD	**26**	12	11	**26**	17	16	**28**	17	**53**	**14**	14	40	28	21	50	20	**50**	**38**	30	50	32	28	121

Universitat Autònoma de Barcelona

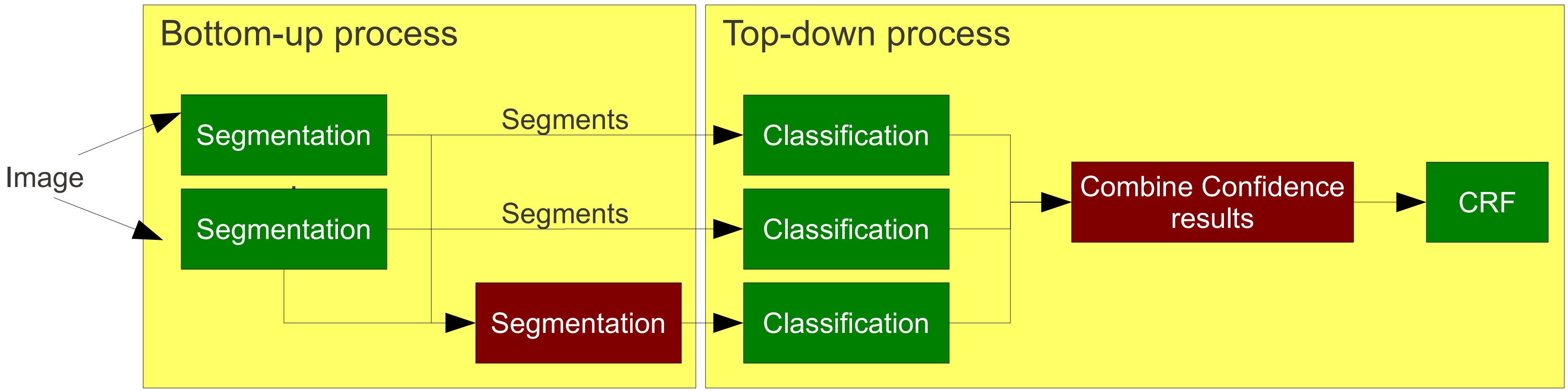
Votation with existing segmentations



	Background	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Dinningtable	Dog	Horse	Motorbike	Person	Pottedplant	Sheep	Sofa	Train	TVmonitor	Avg Accuracy
Votation Best Mix	**48**	20	21	**16**	10	**30**	**32**	42	56	23	19	**35**	51	18	63	**52**	**28**	20	29	40	34	**33**
Votation Worst Mix	34	14	**22**	5	**18**	25	28	21	55	25	17	25	30	**28**	54	34	36	**21**	30	38	33	28
Pantofaru et al. (2008) Best Mix	38	**30**	21	11	14	9	25	24	**61**	**36**	**21**	14	**53**	24	**65**	**52**	27	15	**31**	**42**	**42**	31
Pantofaru et al. (2008) Worst Mix	38	14	14	14	8	6	**32**	**45**	52	19	12	18	30	24	56	**52**	24	15	24	**42**	32	27

Universitat Autònoma de Barcelona

Combine all together

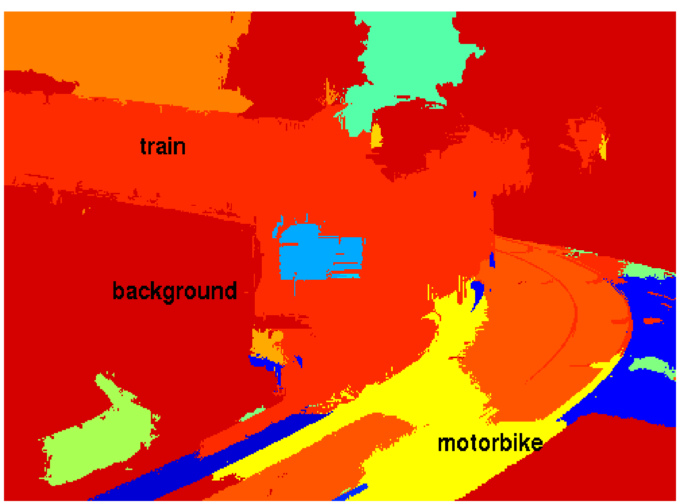


	Background	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Dinningtable	Dog	Horse	Motorbike	Person	Pottedplant	Sheep	Sofa	Train	TVmonitor	Avg Accuracy
Best Mix	49	21	20	10	15	**9**	32	48	56	28	13	**37**	56	19	61	48	33	32	**45**	44	38	**34**
Mix all	50	16	19	15	11	7	32	47	**62**	**29**	**14**	31	53	20	63	58	27	26	39	43	38	33
Viitaniemi (2007)	23	19	21	5	**16**	3	1	**78**	1	3	1	23	**69**	**44**	42	0	**65**	**40**	35	**89**	**71**	30
Fulkerson et al. (2009)	56	26	**29**	**19**	**16**	3	**42**	44	56	23	6	11	62	16	**68**	46	16	10	21	52	40	32
Ladicky et al. (2007)	**78**	6	0	0	0	0	9	5	10	1	2	11	0	6	6	29	2	2	0	11	1	9
Pantofaru et al. (2008)	59	**27**	1	8	2	1	32	14	14	4	8	32	9	24	15	**81**	11	26	1	28	17	20

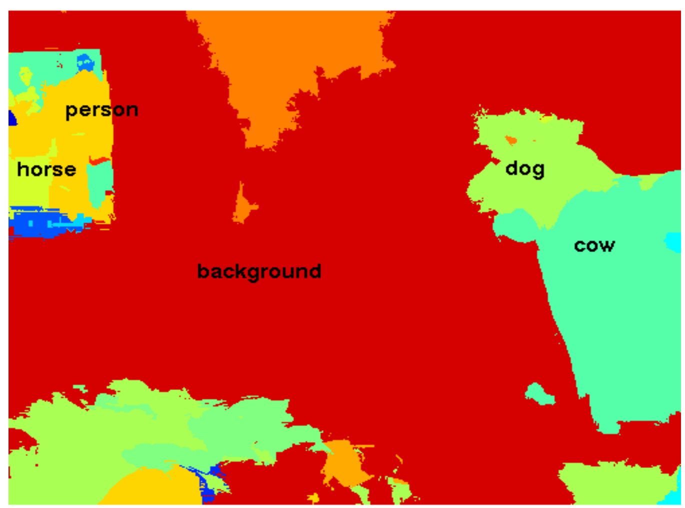
Universitat Autònoma de Barcelona

Method's drawbacks

Votes CRF



Votes CRF



Votes CRF

Conclusions

A novel approach for combining different segmentations to obtain a better segmentation.

A novel approach for combining classification results from several segments to better recognize objects.

Superpixels don't provide the best level of representation of the objects.

Future work

- Adding a weight to each segmentation.
- Create a better RAD segmentation.
  - Investigate in adding spatial locality to the dominant colors.
- Psychophysical experiments to evaluate our approach perceptually
  - Incorporate image level priors.
  - Use object detection to guide segmentation.
  - Use other way for segments goodness measure.

Thank you

RAD Segmentation undesired behavior

- RAD fails in the absence of a large variation in colours in one of the images.

RAD Segmentation undesired behavior

- RAD fails in the absence of a large variation in colours in one of the images.

RAD, Segs = 1

Segments neighborhood

- Introduced by Fulkerson et al.(2009)
- Approach:
  - Extract superpixels from the image (Oversegmentation)
  - Map the image as a Graph G(S, E).
  - Sum histograms from each Superpixel (S) with others less than N nodes away in the graph.

$$H_i^N = \sum_{s_j | D(s_i, s_j) \leq N} H_j^0$$

Our framework

Our framework is divided into five main parts

- Generating multiple segmentations.
- Combining segmentations to get a better new segmentation.
- Describing and classifying regions.
- Combine classification results using the votation technique.
- Refining final results with a CRF.

Our framework is based on the VLBLOCKS by Fulkerson et al. (2009)

Describing and classifying regions

We used the exact parameters in the VLBlocks framework by Fulkerson et al. (2009)

It extracts SIFT features from every pixel.

It uses bag of features framework for classification.

It quantizes our descriptor using K-means and aggregates them into an L1 histogram.

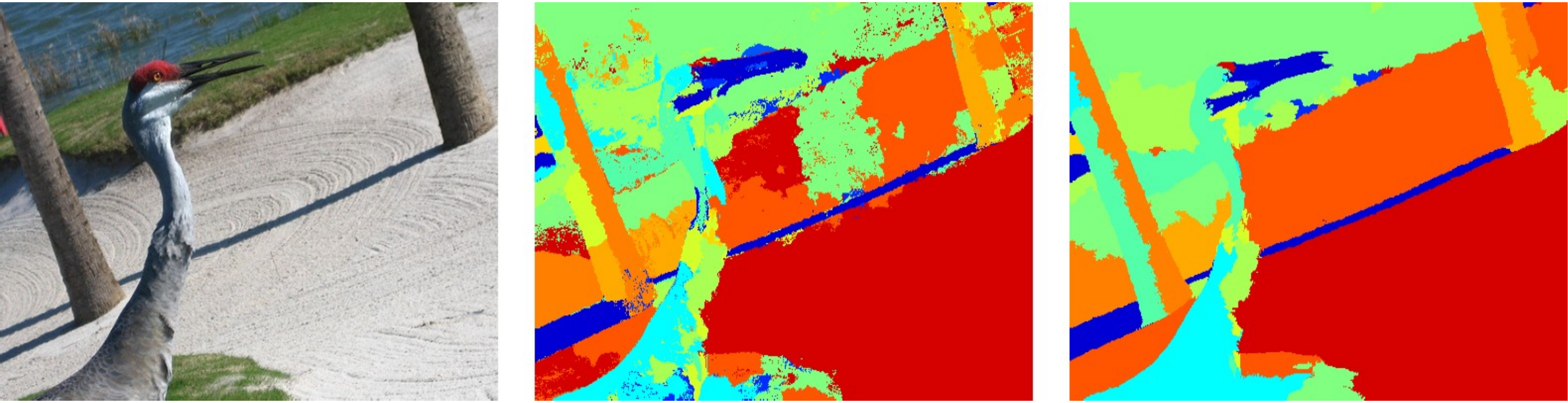
It trains one vs rest SVM with an RBF of chi squared.

Refining classifications with a CRF

We used the same framework from the VLBlocks framework by Fulkerson et al. (2009)

They need the results from the SVM to construct the unary potentials.

They need the difference of neighboring segments colors in the LUV colorspace to identify the pairwise edge potentials.



Introduction

- We present an approach for enhancing image segmentation and object localization by combining segments from different segmentation algorithms.

- Image segmentation?
- Object localization?