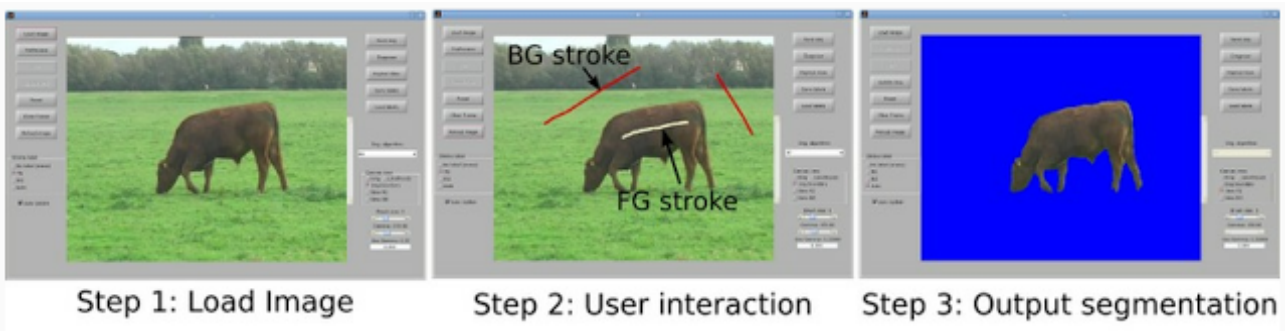
5th July 2012 Datasets for interactive segmentation

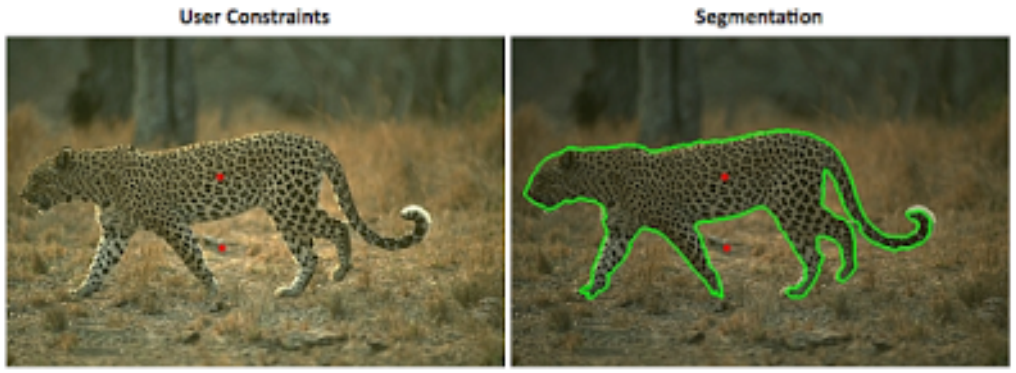
We are currently preparing a paper related to interactive segmentation and, as in most papers, we were faced with the problem of selecting a dataset to segment from. The task has not been as easy as it seemed, so I would like to share my conclusions in case any other researcher has similar needs. I must warn that, in this post, I have linked to the websites containing the datasets. All of them include references to the associated scientific publications, so you can refer them in your papers if you decide to use them.

In our work, we consider the ***interactive segmentation*** goal is to assist a user into an accurate extraction of an object from a background. A former author in this blog, Neus [http://bitsearch.blogspot.com.es/search/label/Neus] , already defined the problem [http://bitsearch.blogspot.com.es/2010/07/improving-usability-of-gat-comparative.html] as well as how to evaluate [http://bitsearch.blogspot.com.es/2010/08/methodology-for-evaluating-interactive.html] the possible solutions. The following figure, from Gulshan et al [http://www.robots.ox.ac.uk/~vgg/software/iseg/] , provides an example of the classic workflow of such systems.



[http://1.bp.blogspot.com/-gwih-GadBDs/T_TVWbmwlJI/AAAAAAAABVQ/uZNa12i_wrU/s1600/2010-Gulshan-GUI.png]

A first and extended dataset in terms of segmentation (not interactive) is the Berkeley Segmentation Dataset (BSDS) [http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/] published in 2000 by Fowlkes, Tai and Malik [http://dx.doi.org/10.1109/ICCV.2001.937655] , and currently mantained by Pablo Arbelaez [http://www.cs.berkeley.edu/~arbelaez/] , Charless Fowlkes [http://www.ics.uci.edu/~fowlkes/] and David Martin. This collection originally contained 300 images that was later extended with 200 additional ones. These images were extracted from the Corel dataset and were manually segmented by human users. As a result, the boundaries of the regions that compose every image are available, following a human criteria. This dataset though does not provide any semantic interpretation of the contents, so there is no discrimination between what regions correspond to objects and which to the background. In a posterior work [http://www.cs.berkeley.edu/~arbelaez/constrained.html] , Arbelaez and Cohen (2008) also used the term ***interactive segmentation*** in that sense, using the user constraints to determine the contours in the image, but with no semantic interpretation.



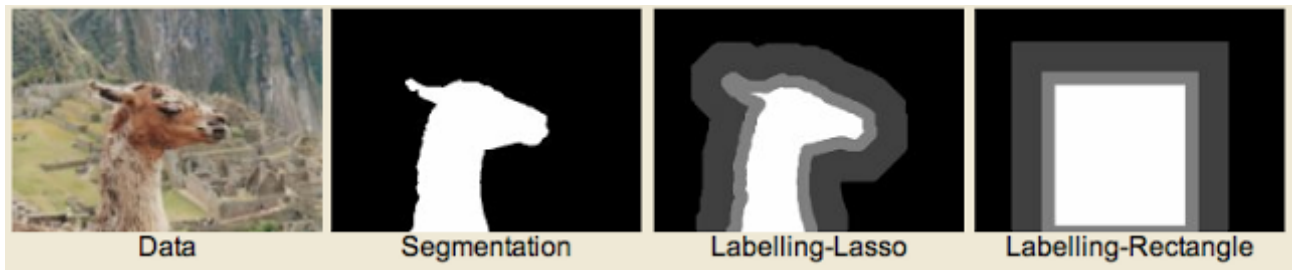
[http://3.bp.blogspot.com/-GFXdL3bjSfw/T_TKCk-Ftil/AAAAAAAABUc/rk56GxI1BGw/s1600/2008-Arbelaez.png]

[http://1.bp.blogspot.com/-YVcWOa62qWM/T_TLQ3M5wiI/AAAAAAAABUk/ICEh5OTgUy0/s1600/2009-McGuinness.png]

Given that our interpretation of interactive segmentation aims at extracting an object, the BSDS lacks this type of ground truth over which our results could be compared to. This missing part was solved by the work of Kevin McGuiness [http://elm.eeng.dcu.ie/~mcguinne/] and Noel O'Connor [http://elm.eeng.dcu.ie/~oconnorn/] from Dublin City University (2009), who manually generated accurate object-background masks [http://kspace.cdvp.dcu.ie/public/interactive-segmentation/ground-truth.html] . They did not process the whole BSDS dataset, but extracted 100 objects out of 96 images; so some images included more than one object. The masks were manually generated with the GNU Image Manipulation Program (GIMP) [http://www.gimp.org/] .

The BSDS was also taken as a starting point for one of the most influencing works in terms of interactive segmentation: the GrabCut [http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm] by C. Rother, V. Kolmogorov, A. Blake and M.Brown at Microsoft (2004). They built a dataset of 50 images, 20 of which come from the BSDS. In addition to the object-background mask, the dataset contains labelling tri-maps that imitate the output that would be obtained when using a lasso/pen tool or ractangular marker. Unfortunately, even the best works are not error resilent proof, and when referring to the list of BSDS that are to be used, there is a mistake: the referred ID 124080

does not exist...



[http://2.bp.blogspot.com/-702_XBpW-SY/T_TORfgRVoI/AAAAAAAABUw/bgEHNR2-nz8/s1600/2004-GrabCut.png]

The problem of the missing image can be partially solved by looking to an additional set of data provided from the GrabCut webpage [http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm] : a list of bounding box coordinates [http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/DATA/ICCV09_new_bounding_boxes.zip] coming from a work by Lempitsky, Kohli, Rother and Sharp (ICCV 2009). In this file the non-existing 124080 becomes 124084, which is actually part of the BSDS dataset. These bounding box coordinates are also very valuable for our work, as we propose an algorithm to use this prior to extract the object.

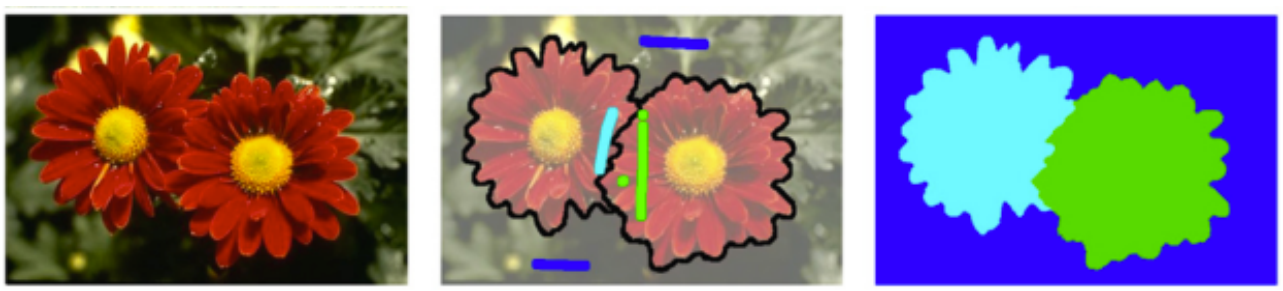
[http://1.bp.blogspot.com/-gnxdp-aT2SE/T_TPoObL19I/AAAAAAAABU4/RcHfuQuG7zM/s1600/2009-Lempitsky.png]

[http://3.bp.blogspot.com/-sIMAXXPiikU/T_TRmR9pHRI/AAAAAAAABVA/Gci7YF6cgAw/s1600/2010-Gulshan.png]

In addition to the rectangular markers (bounding box), our work also deals with solutions that use scribbles as markers. We found a very complete dataset

with brush strokes [http://www.robots.ox.ac.uk/~vgg/data/iseg/index.shtml] (scribbles) published by V. Gulshan, C. Rother, A. Criminisi, A. Blake and A. Zissermann from Oxford (CVPR 2010). This dataset uses 49 images from GrabCut, in addition to 102 from other datasets. If you are a careful reader, you must have noticed that the quantities are not matching, because GrabCut actually contains 50... Well, the problem, as you must have guessed, is in the misterious 124084 image, which was not considered in the Oxford dataset. So we could obtain scribbles for all GrabCut images but one. How to solve this issue ?

We fortunately found still another dataset that works with scribbles and GrabCut, the one associated to the Deformable Graphs [http://structuralsegm.sourceforge.net/] proposed by Noma, Graciano, Cesar, Consularo and Bloch (2012). They provide a Java implementation of their segmentation algorithm together with SOME scribbles. Yes, you have read well: SOME. Their experiments used some images from here and there. Six from the Berkley sub-set in GraphCut and, fortunately, the misterious 124084 was among them. So we can complete the collection ! The next figure shows the image and the segmentation obtained with the provided scribbles and algorithm.



[http://3.bp.blogspot.com/-e2fnbHINVBE/T_TTPPtPpxI/AAAAAAAABVI/2Vdqic4-I5Y/s1600/2012-Noma.png]

[http://3.bp.blogspot.com/-VBSjH2md9TE/T_TWpCsmuFI/AAAAAAAABVY/uMWmFDM-a_M/s1600/robot-painter-nanotech.jpg]

The provided bounding boxes and scribbles will allow us to run the experiments we are aiming at for the current work. However, there still exist a more refinate approach that allows a more accurate analysis of the behaviour of the interactive segmentation algorithm. These type of algorithms are expected to generate the output after multiple interactions from the user-side. After each interaction, the system will suggest a mask that may trigger a new interaction from the user. These partial interactions could be generated by a **robot user** and the evolution of the

segmentations be tracked. Both works from Gulshan et al [http://www.robots.ox.ac.uk/~vgg/software/iseg/] (Oxford) and McGuinnes and O'Connor [http://kspace.cdvp.dcu.ie/public/segmentation-evaluation/downloads.html] (Dublin City University) offer this possibility and, in the later case, the authors also provide an interface to visualize the results, by plotting the accuracy vs # interactions curves.

To sum up, using scientific datasets is not always that easy. While using a public dataset facilitates the comparison and repeatibility of the results, finding the right one requires by itself an intensive research. Even once found, technical issues appear that sometimes require some extra time that was not expected. This conclusions are similar to the ones I reached last year when using a single dataset, the ETHZ Shape, which posed by itself several challenges [http://bitsearch.blogspot.com/2011/08/adaptation-of-ethz-shape-database-to.html] .

Posted 5th July 2012 by Xavi Giró-i-Nieto

Location: Universitat de Colúmbia, 116th St, Broadway, Nova York 10027, Estats Units d'Amèrica

Labels: evaluation, interactive segmentation, xavi