

10:00 am – 10:30 am – OS29 - Nov 08, Friday – 17th Oral Session

ON THE USE OF HAUSDORFF DISTANCE FOR EVALUATION OF BREAST IMAGE SEGMENTATION

R. S. Marques,
A. Conci,
Federal Fluminense
University

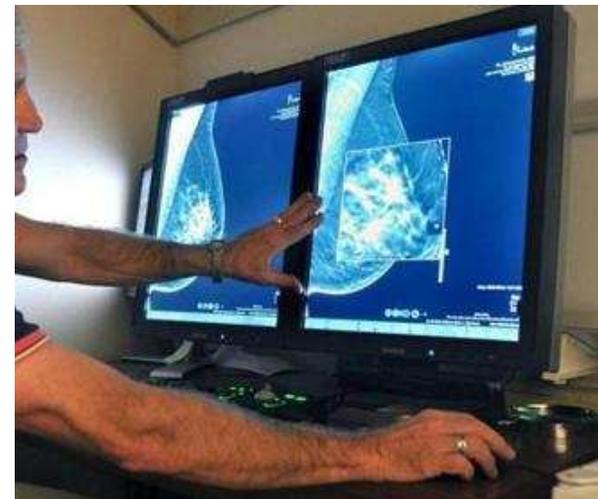
T. MacHenry,
York
University



Introduction

Image Processing (IP) and **Pattern Recognition (PR)** techniques **have been applied in analysis** of the most **commons types of medical diagnosis.**

This work is related to diagnosis of **breast** diseases.

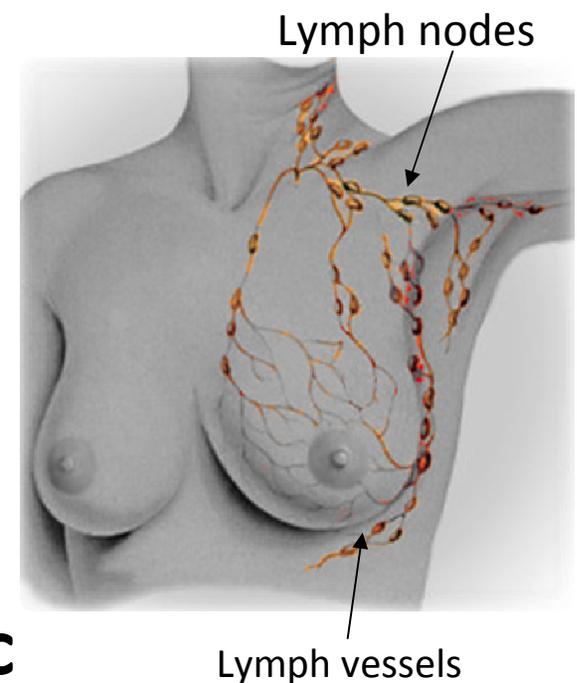


Region of interest (ROI).

In **IP** and **PR** the first step, after acquiring the image, is to **separate the important elements in it**, that is, to obtain the **region of interest (ROI)**.

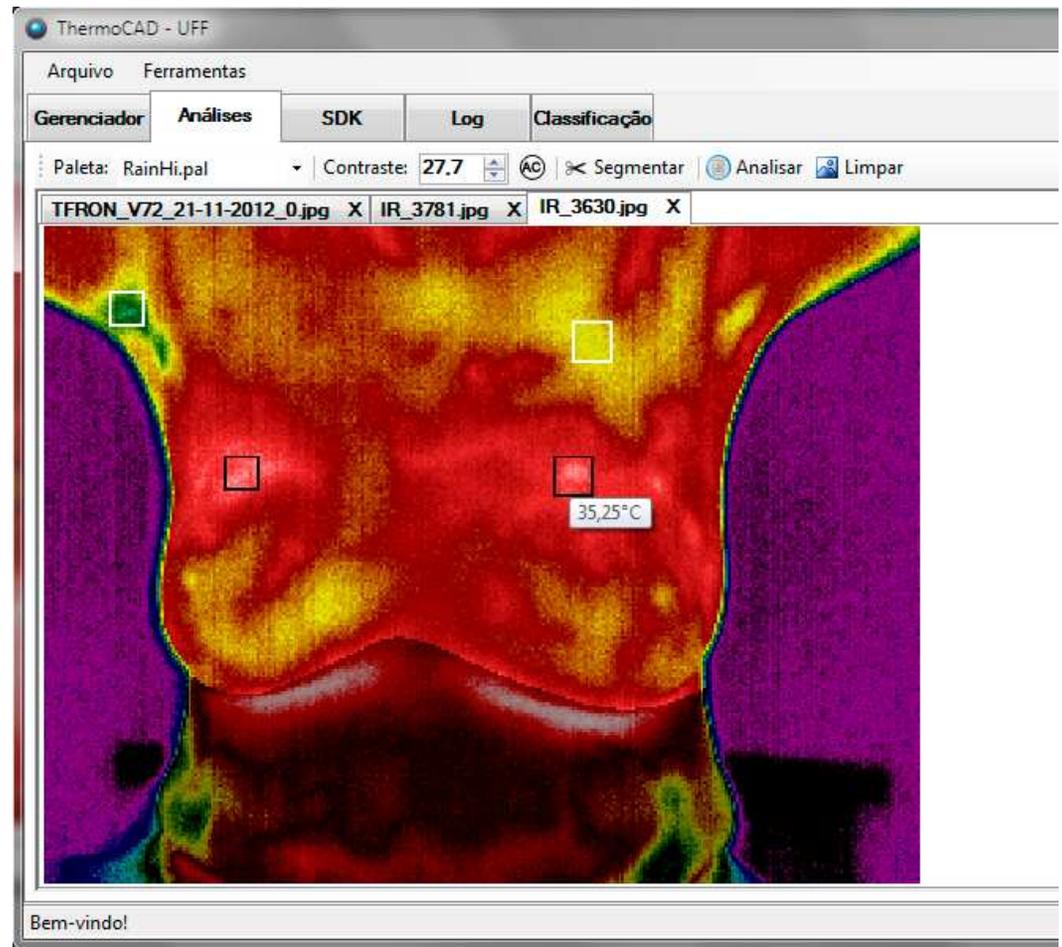
Approaches for this must be **compared by using public available databases**.

For **Infra Red (IR)** images the **ROI** must include the **axillar lymphatic nodes**.



In computer aided diagnosis (CADx) and detection (CADE) systems ROI identification algorithms must be:

- A complete **automatic procedure** (found with no user interaction);
- Segmented **in real time**, that is during the patient examination; and
- Must be quantitatively **evaluated** by using a **ground truth (GT)**.

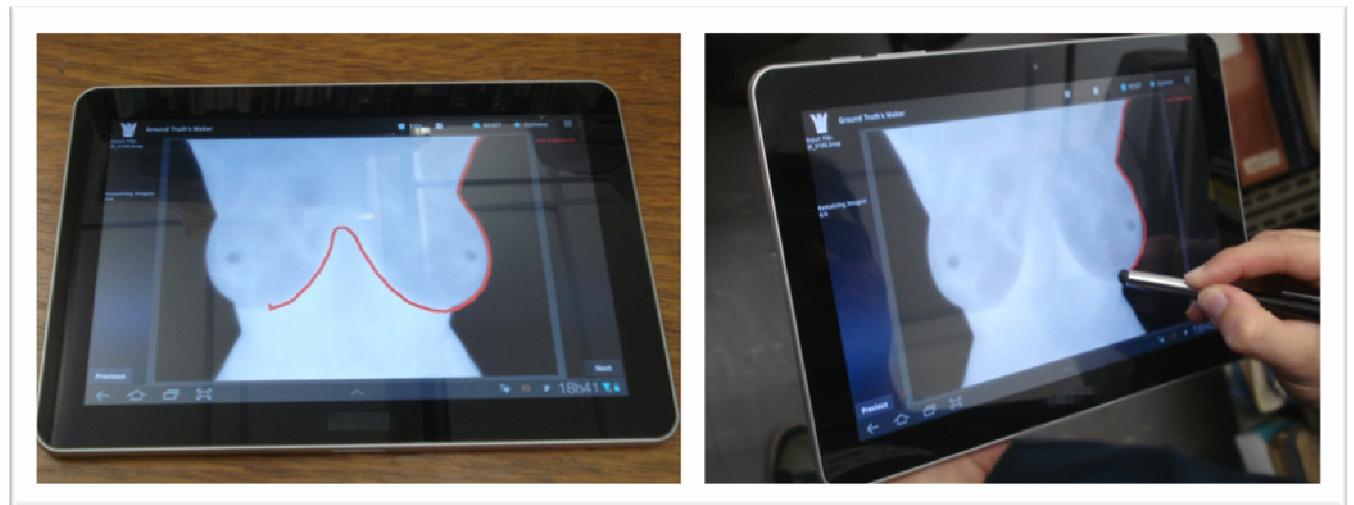
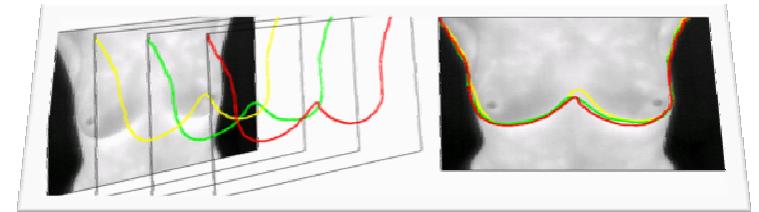
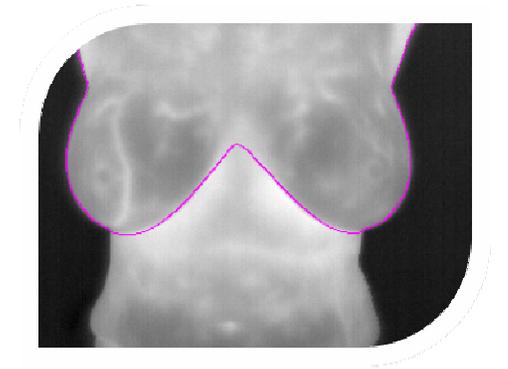


Ground truth (GT) for IR images

Correct segmentation by a number of experts in the area must be considered.

Must present a number of cases with **segmentations made manually by a group of experts.**

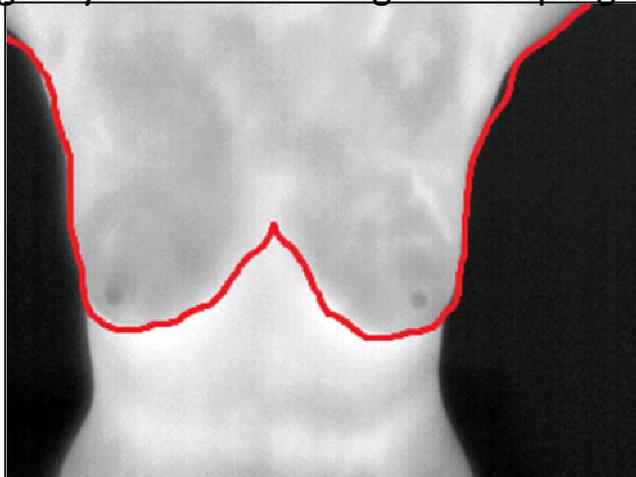
To motivate specialists to spend their time doing this type of work we developed a specific **system** with a number of functionalities and facilities: **GTmaker.**



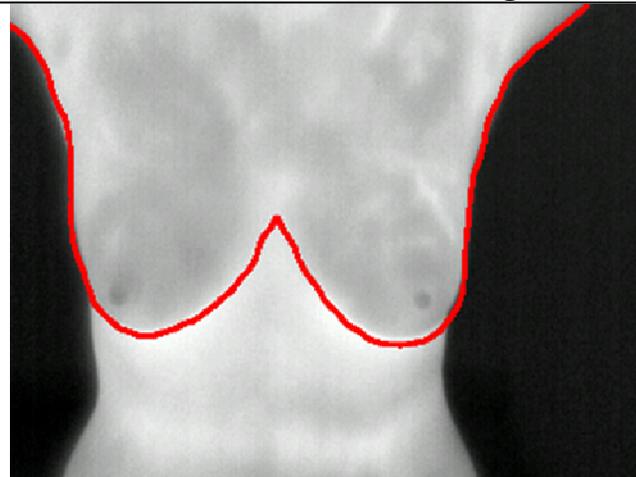
GTmaker:

- This helps **manual segmentation** ;
- Uses **touch screen** technology ;
- Helps in the **evaluation, validation** and **comparison of different techniques** for automatic segmentation of ROI ;
- Is available for **mobile** devices, providing greater mobility and time optimization for the specialists (Android) ;
- Provides: **Curve approximation and interpolations**, “do”, “undo”, “redo”, storage, retrieval, merge, combination of view, etc.

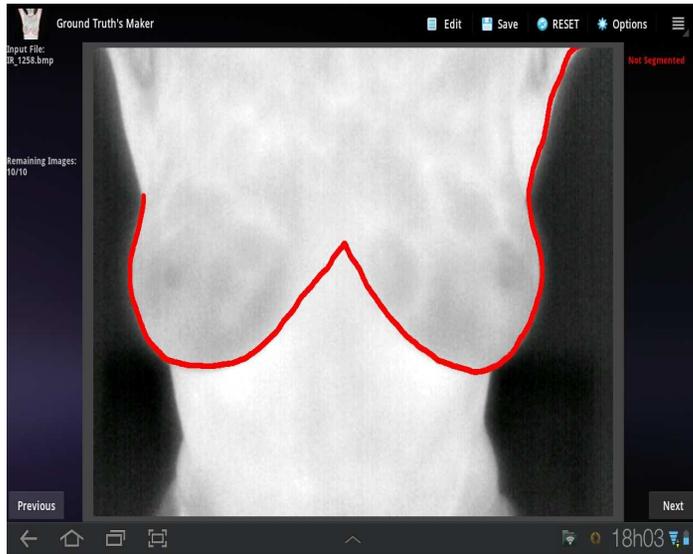
Using only mouse and image editor programs



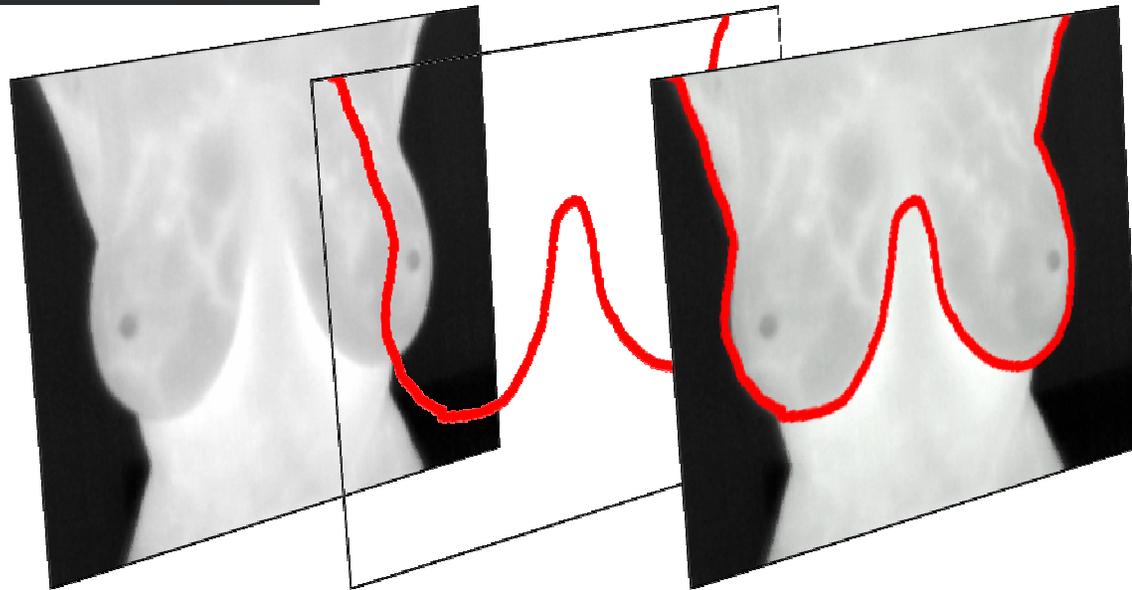
Using GTmaker



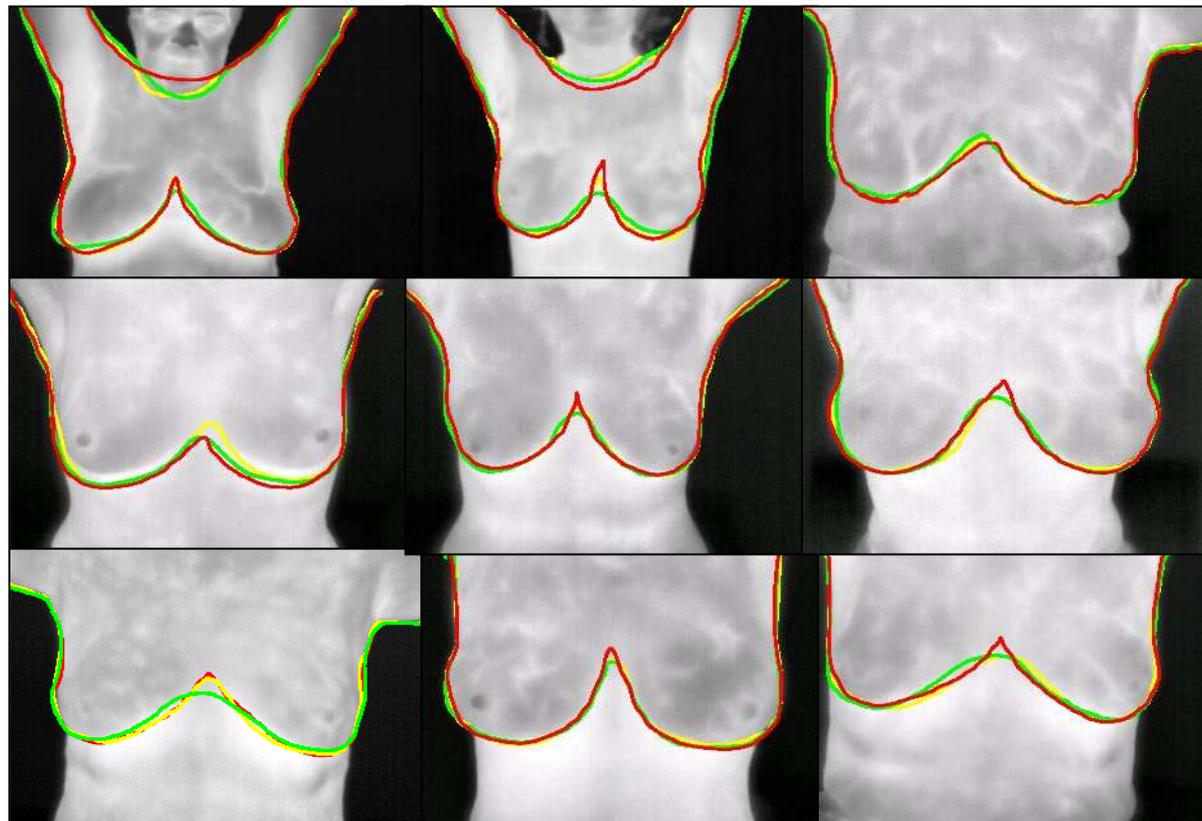
Acquisition and examinations <http://visual.ic.uff.br/en/proeng/software.php>



Sample of some functionalities:
retrieval, merge, combination of view and
Smoothing.



Ground Truth creation: Segmentation by at least three **Medical Doctor** of the **Radiology Service** of the University Hospital of the Fluminense Federal University (UFF), Niteroi, Rio de Janeiro (Brazil) is used to form each image of the GT



How the **ground truth** is created by combining ROI of experts :

285 patients
285 x 3 = 855 images

Public
thermo mammographic
databases

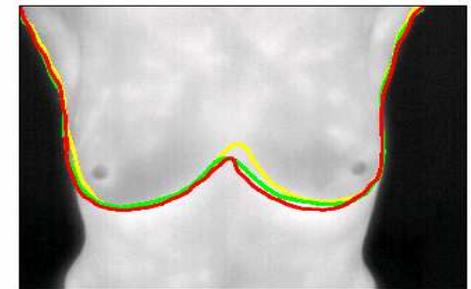
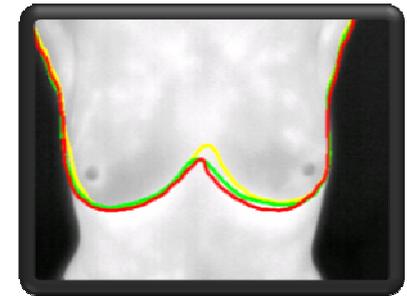
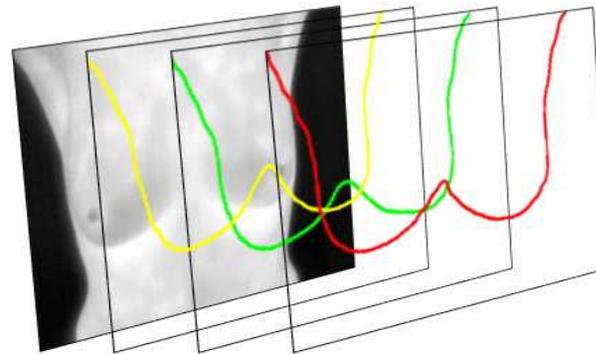
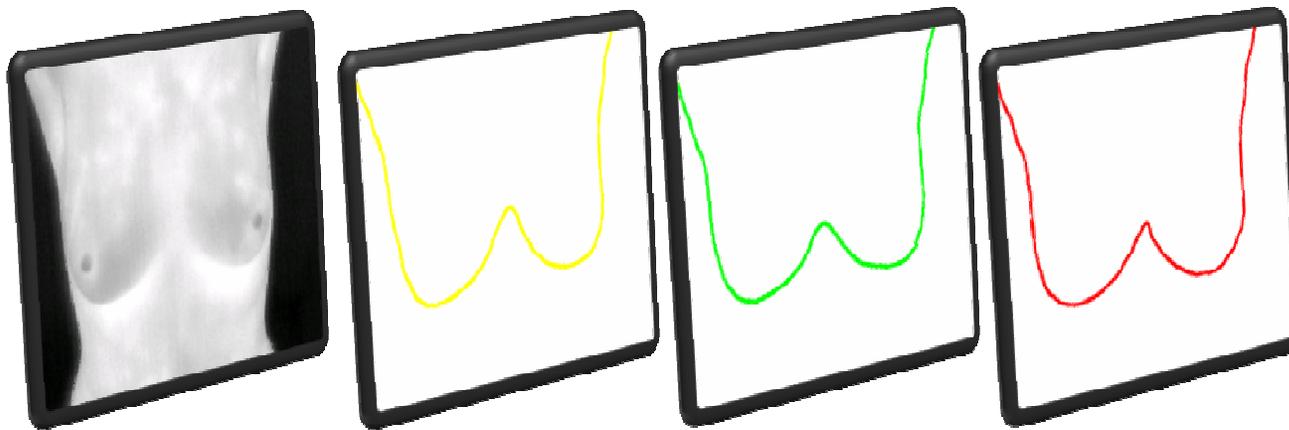
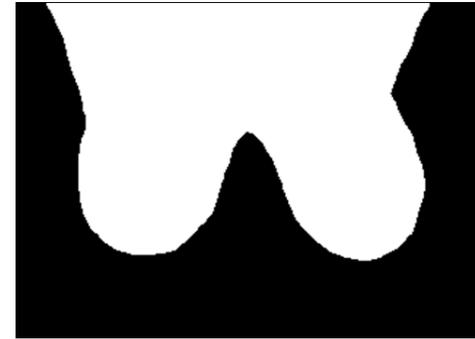


image IR_8269 of database



The fourth **ground truth** , or the unified **ground truth**



- In addition of the three ground truths generated for each image, a fourth ground truth was generated based on a **voting policy**.

According to this voting policy a pixel of coordinates (x,y) will belong to ground truth if at least **K manual segmentations define that pixel as object (ROI)**,

where

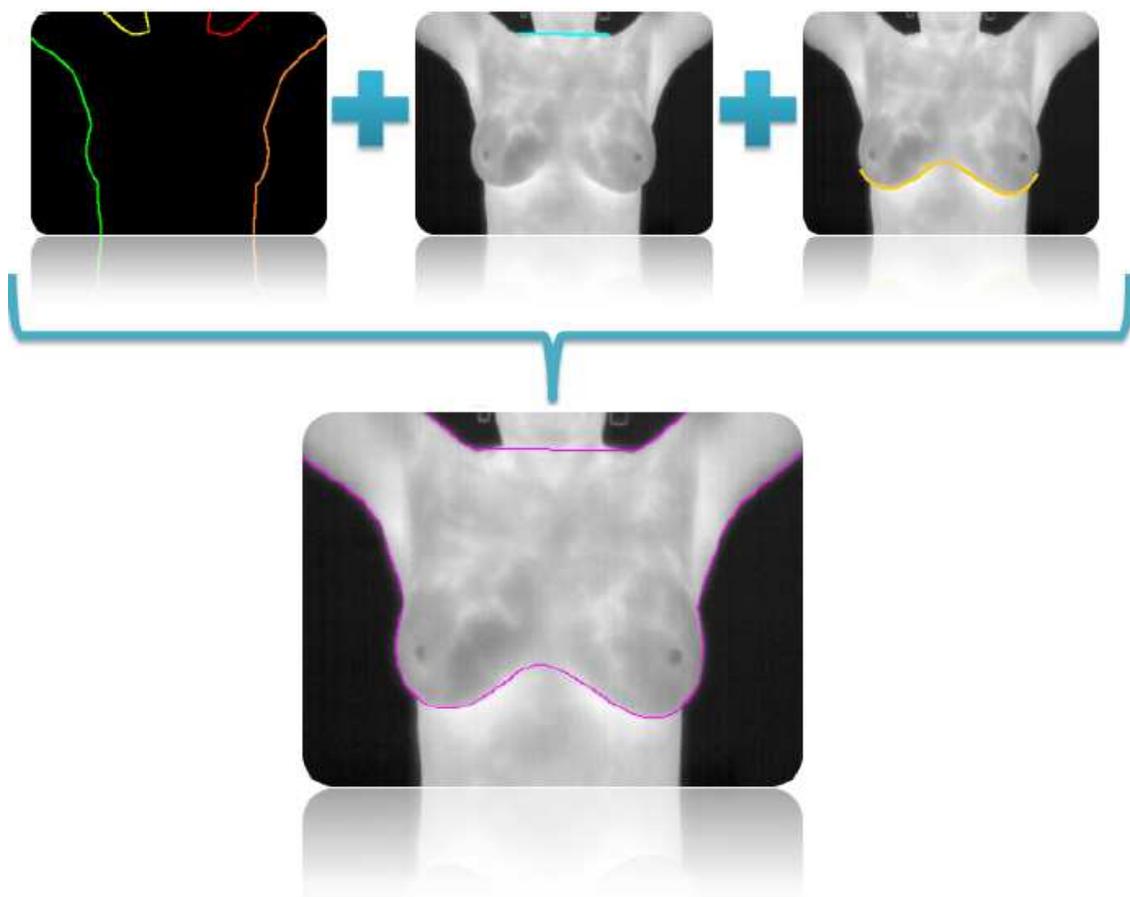
$$K = (j+1)/2$$

and j is the number of manual segmentations .

New algorithm for automatic ROI segmentation

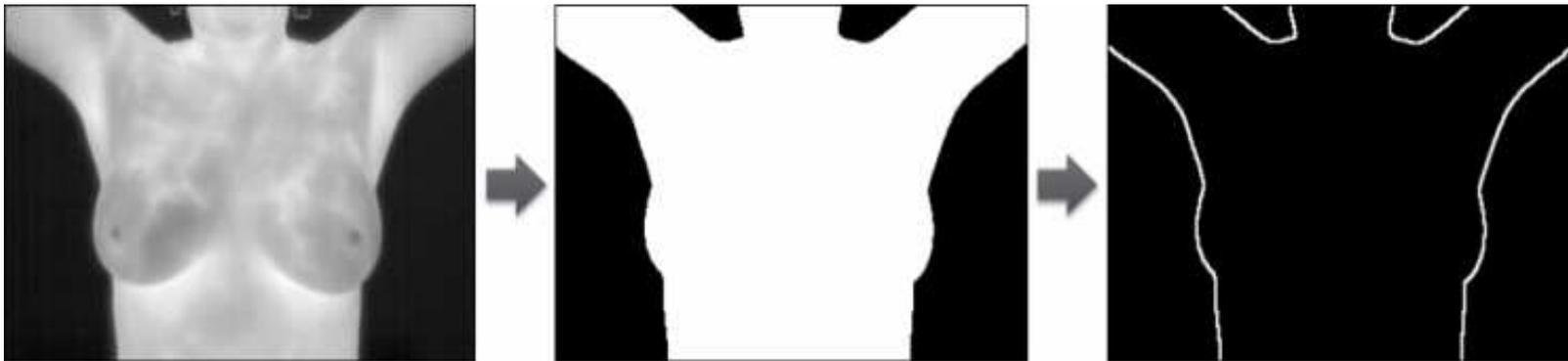
Our automatic segmentation of the region of interest (ROI) can be divided into **three main steps**:

- detection of the lateral,
- upper and
- lower boundaries



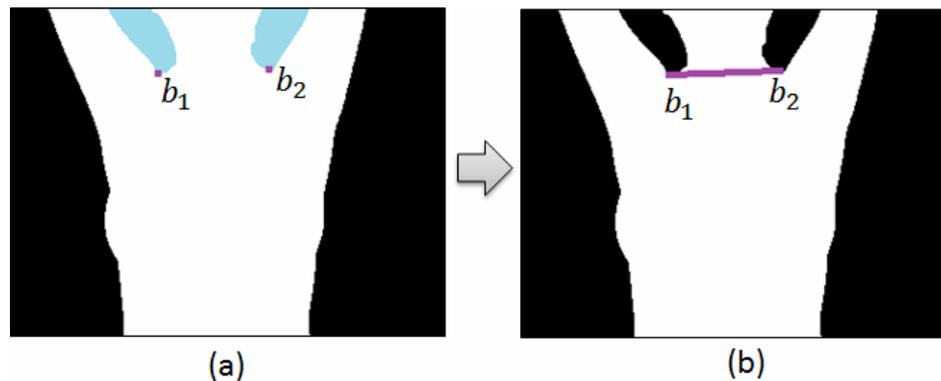
The lateral limits of the ROI are patients' body borders.

- For this identification: First, the patient **body is separated from background** by using **Otsu's thresholding method**, then a **Laplacian filter** is applied to obtain the edges



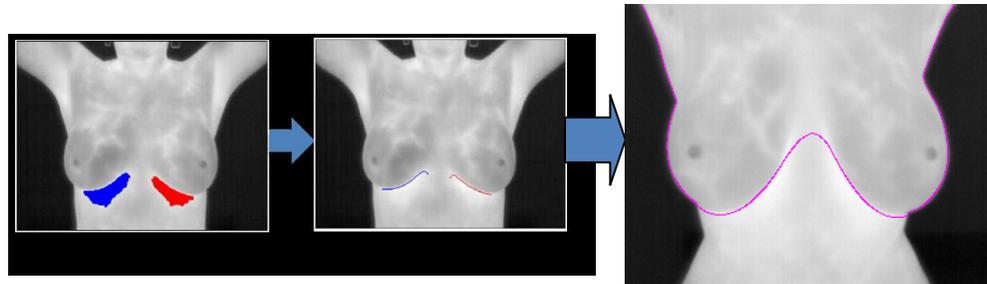
The upper bounds are found through the detection of shoulders or raised arms.

Algorithm finds the presence of arms. The lower points of the background area identified and used to separate the neck from the rest of the ROI



Lower bounds:

- Are focused on the **detection of the inframammary fold**, characterized by **high temperatures**.
- Curves are defined using: **Uniform Quadratic Spline** for the **left and right** curves joined by **Spline interpolation** forming a unique curve.
- ROI can then be isolated from the rest of the image.



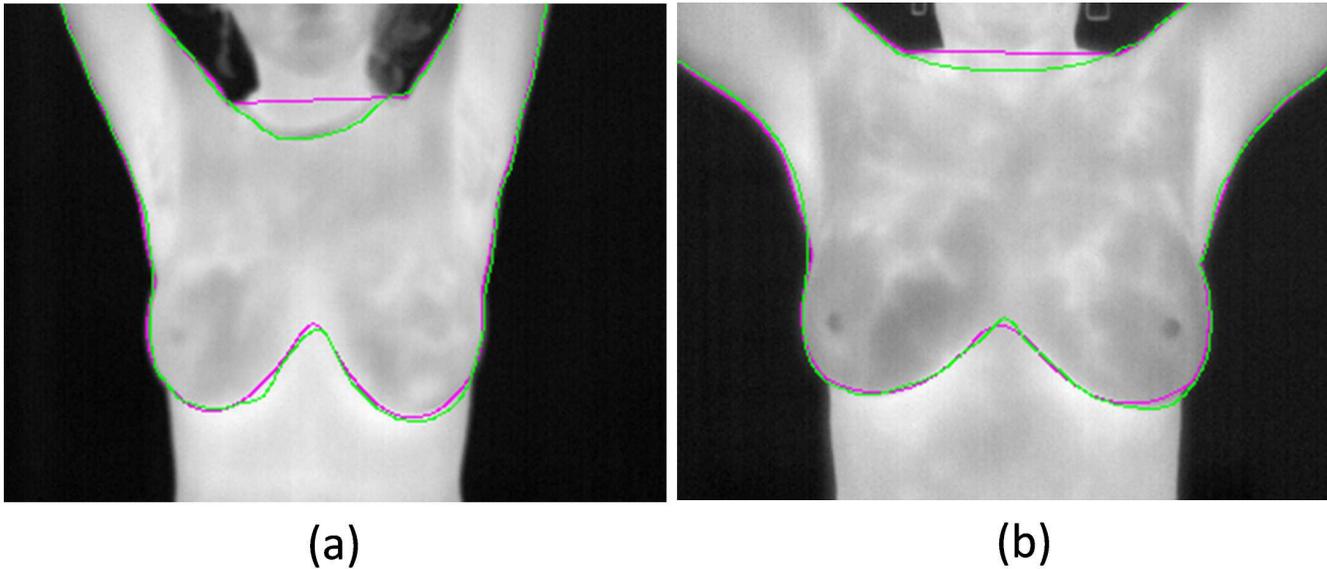
Details:

- All algorithms have been implemented in **C++** computer language using **OpenCV** libraries version 2.3.1 , and **GSL** (GNU Scientific Library) version 1.8.
- The **original images** and their **ground truths** can be viewed or downloaded at <http://visual.ic.uff.br/en/proeng>.
- They are available as a part of a research approved by the **Ethics in Research Committee** (CEP) of HUAP/UFF on 04/06/2012, registered in Brazilian Ministry of Health Platform under the number (CAAE) 01042812.0.0000.5243 for **early detection of breast diseases**.

How to compare ROI segmentations?

Intersection of surfaces

Hausdorff distance (HD) (provides a **unique measure** which is **intuitive** and **simpler** but uses considerably more computational **time**).

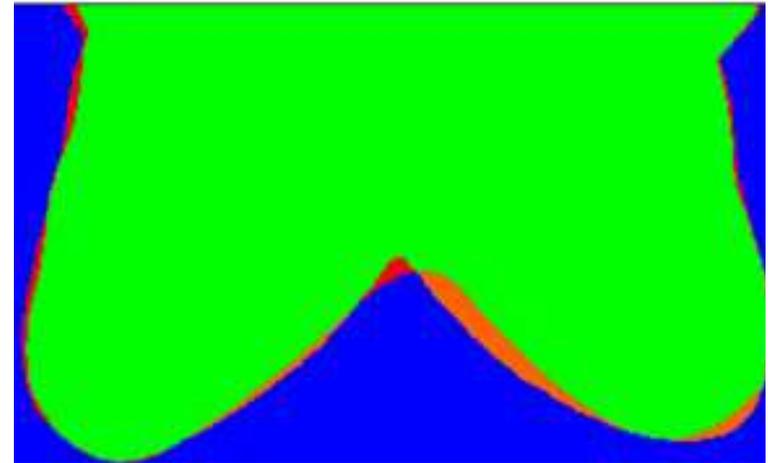


Overlay of edges from **proposed algorithm** (**pink**) and the **ground truth** (**green**)

(a) IR 0438, HD=20.52 and (b) IR 0861, HD= 11.31

Intersection of surfaces

Consider *TP* (True Positive); *TN* (True Negative); *FP* (False Positive) and *FN* (False Negative) by the green, blue, red and orange, area respectively.



Accuracy (**ACC**): $ACC = (TP + TN) / (FN + FP + TN + TP)$

Sensitivity (**SEN**): $SEN = TP / (TP + FN)$

Specificity (**ESP**): $ESP = TN / (TN + FP)$

Predictive positive (**PDP**): $PDP = TP / (TP + FP)$

Predictive negative (**PDN**): $PDN = TN / (TN + FN)$

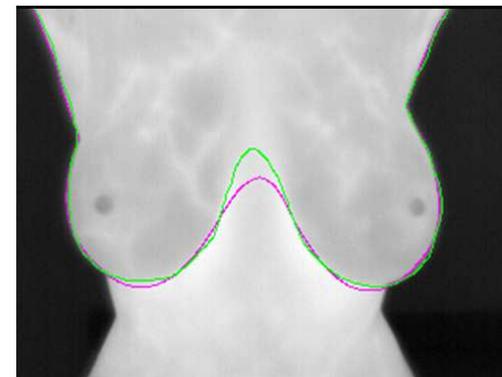
Comparison between the **proposed segmentation** versus **ground truth** using **HD** and last equations for (Intersection of surfaces) some images of database.

IR	HD	ACC	SEN	ESP	PDP	PDN
0100	12.76	0.981	0.993	0.973	0.969	0.993
0149	5.00	0.990	0.990	0.989	0.993	0.984
0756	15.56	0.981	0.999	0.972	0.949	0.999
0973	16.97	0.976	0.993	0.961	0.957	0.994
0990	7.07	0.990	0.990	0.990	0.987	0.992
3416	16.28	0.977	0.997	0.967	0.938	0.999
3743	7.28	0.983	0.984	0.980	0.984	0.981
3748	7.81	0.983	0.987	0.979	0.981	0.986

Note: Much simpler and more intuitive using HD !!

An example:

Result of a spline based automatic segmentation (red) for the **IR 0100** overlapped with the ground truth (green).



Improving the time for Hausdorff distance (HD) computation

- HD evaluation by using traditional tools needs considerably more **time**, because all points in each set are considered when computing its distance from each set to the other, followed by an assignment of the greater value to the HD.
- We propose a **new algorithm** for computing the **Hausdorff distance** between **discrete curves** and areas with **no gaps**, in which only the border of the curves is considered.
- This is very adequate for evaluation of ROI distances because they satisfy these conditions.

Hausdorff distance (HD) definition

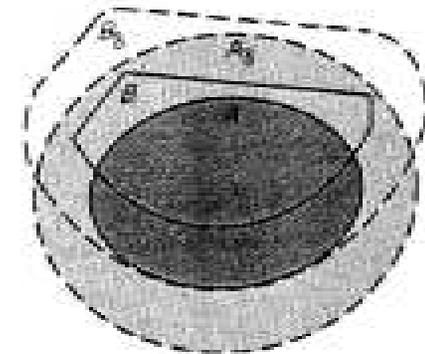
- Let D be a closed subset of R^n and let S denote the class of all non-empty compact subset of D .
- There are some alternative (equivalent) ways of defining *Hausdorff metric* on all non-empty compact subset S , of closed subset D , of R^n .
- Define *the parallel body* of a set A as:

$$A_\delta = \{x \in D : |x - a| \leq \delta \text{ for some } a \in A\}$$

Hausdorff distance (HD) - definition

We make S into a metric space by defining the distance $d(A, B)$ between two sets A, B to be the least δ such that the δ -parallel body of A, A_δ , contains B and the δ -parallel body of B, B_δ , contains A (see image on Figure):

$$d(A, B) = \inf\{\delta : A \subset B_\delta \text{ and } B \subset A_\delta\},$$



Then the two sets A, B is now subsets of the metric space S . Considering the definition of the closure of A and B , denoted by A^- and B^- we have

$$d(A^-, B^-) = \inf\{\delta : A^- \subset B_\delta^- \text{ and } B^- \subset A_\delta^-\} = \inf\{\delta : A \subset B_\delta \text{ and } B \subset A_\delta\} = d(A, B),$$

Moreover by definition of the boundary of A and $B : \partial A$ and ∂B we have:

$$d(\partial A, \partial B) = d(A^-, B^-) = d(A, B),$$

Hausdorff distance (HD) - definition

It is plausible to consider a discrete version of the Hausdorff distance by replacing A with a discrete version of it, say A_Δ . Allowing a slight abuse of notation, let ∂A_Δ denote the intersection of A_Δ with ∂A . Carefully note that this is not the boundary of A_Δ (in fact, it makes no sense to talk about the boundary of a discrete subset of R^n). The same argument applied to the set B and yields a discrete version B_Δ of it. In this way, we can think of A_Δ and B_Δ as subsets of a discrete version of R^n , say Z_Δ , and so we proceed to compute a discrete version of the *Hausdorff distance* between A_Δ and B_Δ based on following discrete version of

$$d(A_\Delta, B_\Delta) = d(\partial A_\Delta, \partial B_\Delta),$$

where $N_8(p)$ denotes the 8-neighborhood of the discrete point p

$$\partial A_\Delta = \{a_i \in A_\Delta : \exists p_j \in A^c, p_j \in N_8(a_i)\} \text{ and } \partial B_\Delta = \{b_i \in B_\Delta : \exists p_m \in B^c, p_m \in N_8(b_i)\}$$

A^c, B^c denotes the complement of sets A and B

*Using this new version, the time of computing the distance between sets is **greatly reduced** !*

Average execution time in **milliseconds** for images using brute force and the new algorithm

The ROIs of the images in two resolutions (320x240 and 640x480) are considered for evaluation of time.

Time	320x240 8-N	320x240	640x480 8-N	640x480
Traditional algorithm	9.57	37.78	76.51	610.65
Proposed algorithm	0.75	1.49	3.98	20.60

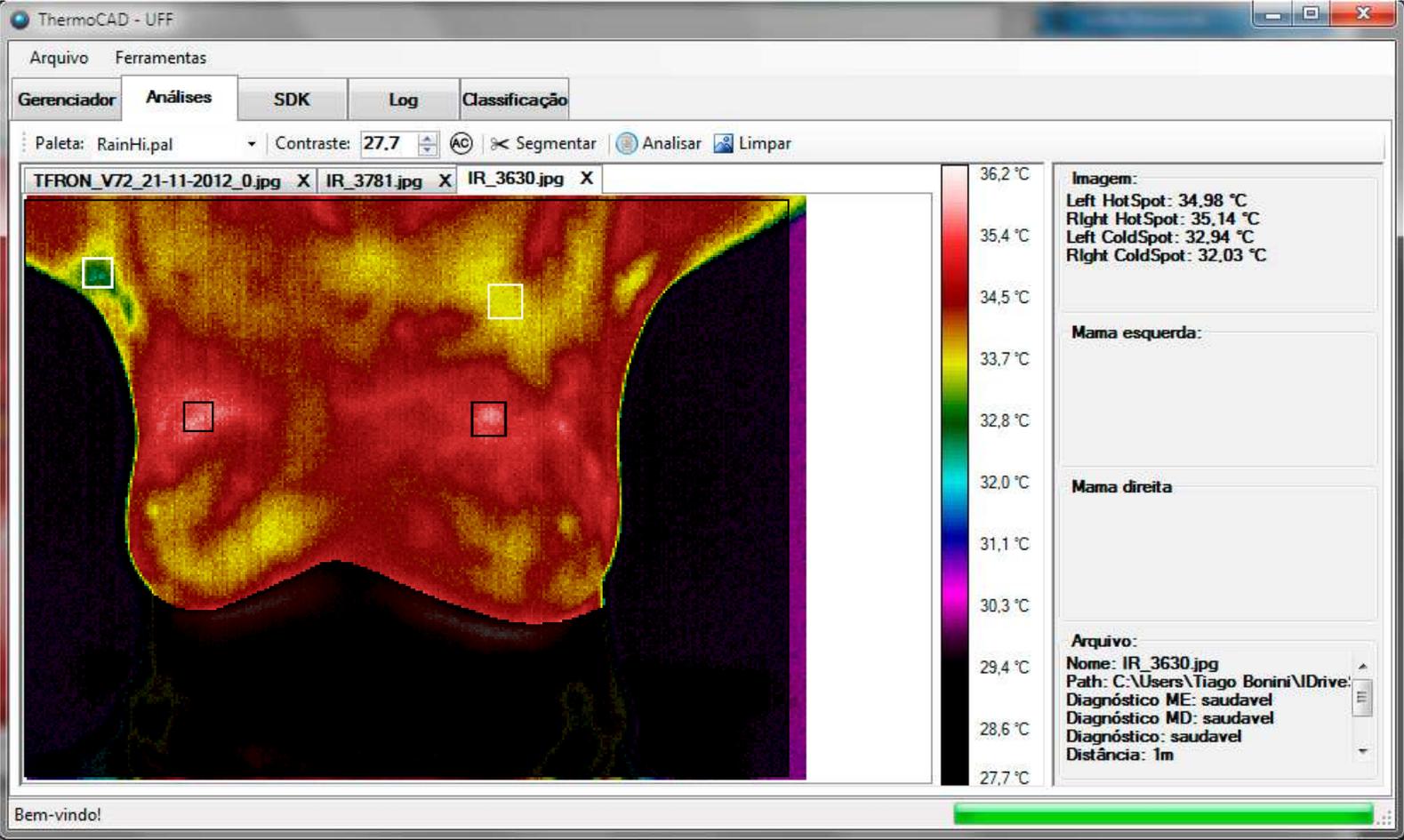
Conclusion

- For CAD systems , segmentation of region of interest (ROI) is an important step, it must be **automatic**, and **subject to validation**.
- This work presents a **new methodology** for the extraction of the region of interest (ROI) of thermographic examinations. It considers its **validation**, as well.
- This is realized through comparison of real thermographic exams available from a **database** being developed by radiologists of the Antonio Pedro University Hospital - HUAP of the Federal Fluminense University and its **ground truth**.
- It is **made from images of volunteers who accepted having published for researches purposes**.

Conclusion (continued)

- To be used on a system for Diagnostic Aid, the **computational time** of the algorithms is a fundamental aspect.
- The efficiency of the methodology is (qualitatively and quantitatively) verified using the **ground truth**.
- To verify the adequacy of the proposed approach for region of interest (ROI) segmentation a new algorithm for evaluation of HD is considered.
- New metrics like the ***Oriented Distance Functions*** proposed by Hanna K. **Jankowski** & Larissa I. **Stanberry** must be considered.

ROI of ThermoCAD-UFF .



Thank you!