Visual quality metrics

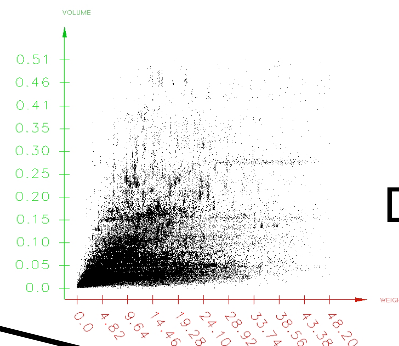
Enrico Bertini and Giuseppe Santucci

Dipartimento di Informatica e Sistemistica

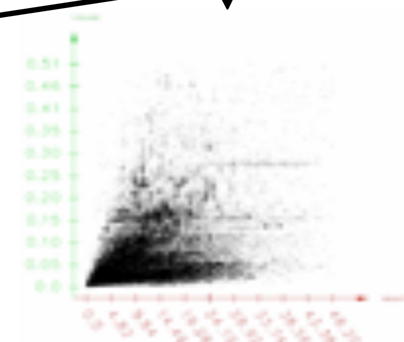
University of Rome ”La Sapienza”

The problem

- Very often data visualizations do not convey data features in a correct way !
- Infovis' enemies:
  - clutter
  - collisions
  - user perception
  - pixel dimension
  - ...



Data



Degraded Visualization

You cannot control what you cannot measure [T. De Marco, 1982]

What are we measuring?

We propose the following taxonomy:

1. Size metrics

2. Visual effectiveness metrics

3. Features preservation metrics

Size metrics

- purpose: basis of any other computation
- examples:
  - number of data items
  - size of pixels
  - ...
- can benefit from perceptual studies (e.g., numerosity), indicating the limits of human perception and thus providing some useful threshold values

- quite intuitive and we do not discuss them anymore (but we use them!)

Visual effectiveness metrics

- purpose: measuring the image degradation, taking into account some disturbing factors
- examples:
  - number of collisions
  - number of outliers
  - ...
- most of the available metrics belong to this class
- we discuss how to **use** these metrics

Features preservation metrics

- purpose: intended for measuring how correctly an image is representing some data characteristics
- examples:
  - Tufte' s lie factor, that is the ratio between the size of an effect, as shown graphically, to its size in the data
  - ???
    - The idea of **comparing** data characteristics against visualization effects, pioneered by Tufte, has not been pursued anymore...
- As a consequence, few proposals are available in this class and we intend to analyze these metrics in detail, discussing how to **define** and **use** them

A four steps methodology for defining feature preservation metrics

1. *Choose the feature to preserve*

2. *Formally define the feature in (1) the data space and (2) in the visualization space*

3. *Validate and tune the definition in the visualization space through user perception*

4. *Define the metric* ***comparing*** *figures coming from data space against figure coming from the visualization space*

- We detail these steps through a 2D scatter plot example

A (not simple) example: step 1

1.Choose the target feature: **density differences**

Therefore our goal is to measure how density differences that **exist** in the data set are presented to the user using a 2D scatter plot.

A (not simple) example: step 2

2.Formally define the feature in (1) the data space and (2) in the visualization space

- to formalize density differences we split the 2D plane in little squares that we call **sample areas** (typical dimension: 0,08x0,08 inches / 8x8 pixels)

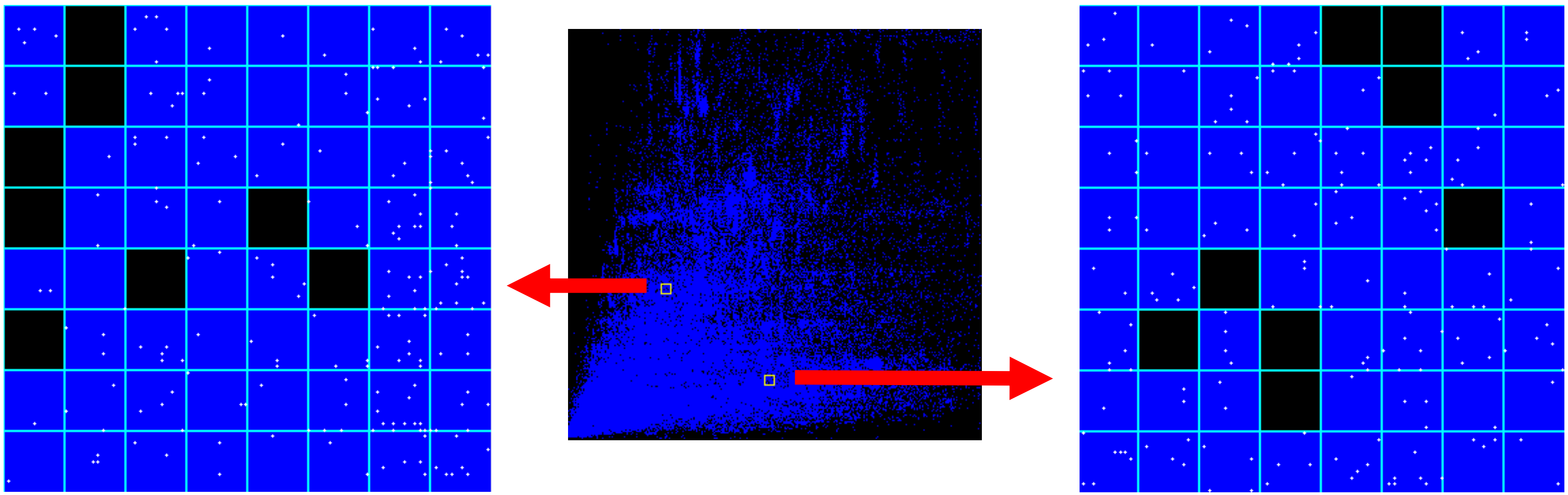
- in the **data space** we measure the **Data Density**, the number of data points within a sample area

- in the **visualization space** we measure the **Represented Density**, the number of active pixels within a sample area

- because of collisions we have that in a sample area:

Represented Density<=Data Density

A (not simple) example: step 2



Data density=205
Rep density=56

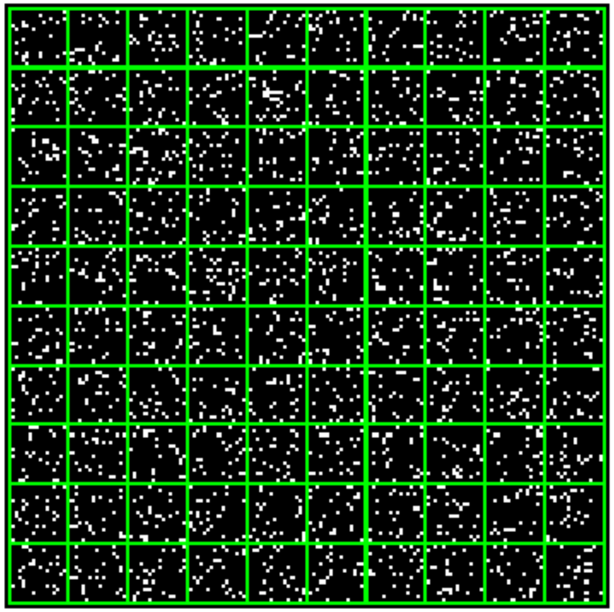
Data density=189
Rep density=56

The actual visualization hides this difference

A (not simple) example: step 3

3. Validate and tune the definition in the visualization space through user perception

the question is: are Represented Density **numerical** differences adequate?



100 sample areas

97 contain 25 pixels

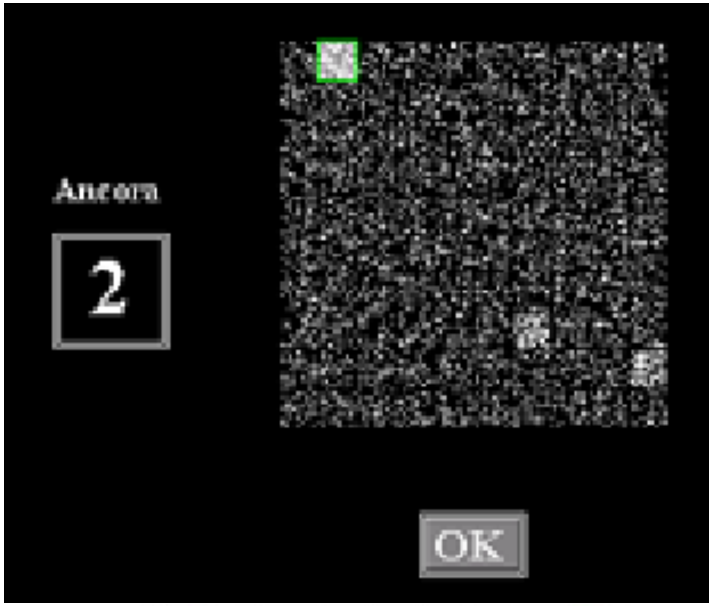
3 contain 38 pixels

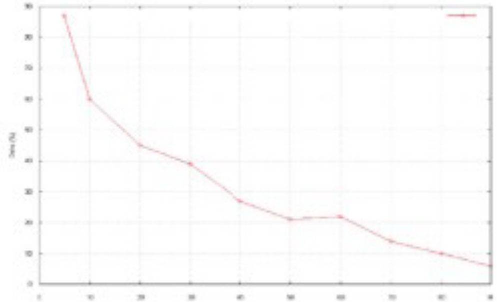
Which ones ?

A (not simple) example: step 3

3. Validate and tune the definition in the visualization space through user perception

- What is the smallest difference in pixels between two sample areas that produces the perception of density difference?
- User test step:
  - image with 100 sample areas
  - 97 with the same number of pixels (basis)
  - 3 filled with extra (delta) pixels
  - the user has to recognize the three more dense areas
- repeated for different basis and deltas





A (not simple) example: step 4

*4.Define the metric **comparing** figures coming from data space against figure coming from the visualization space*

WPLDDr (Weighted *Perceptually* Lost Data Densities ratio)

the metric counts the sample area pairs whose represented density **does not match** the real data density (we weight the comparisons with the involved data points).

$$PDiff(x,y) = \begin{cases} 1 & \text{if } x \geq y + y \times minimum\delta(y) \\ -1 & \text{if } y \geq x + x \times minimum\delta(x) \\ 0 & \text{otherwise} \end{cases}$$

$match(i,j,k,l)$ = true iff

$$PDiff(D_{i,j}, D_{k,l}) = PDiff(RD_{i,j}, RD_{k,l})$$

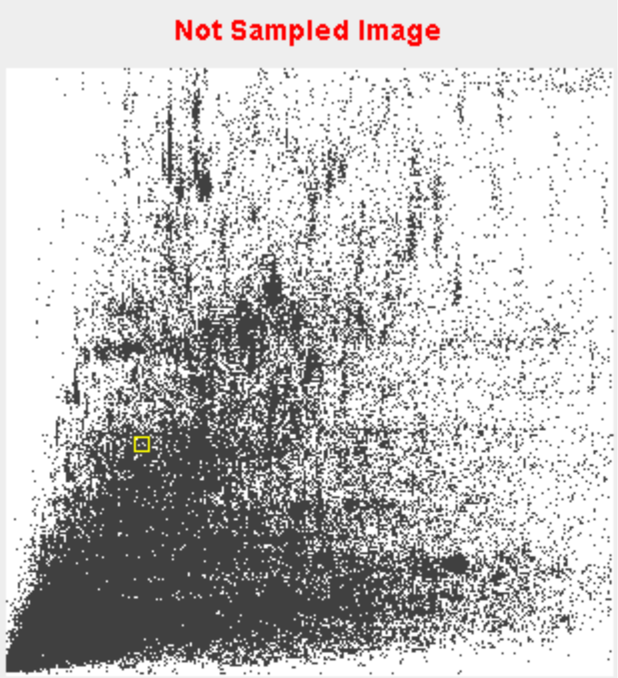
```
function WPLDDr(){
 Let couples=0; \* weighted SA couples
 Let sum=0;     \* weighted non matching SA couples
 foreach distinct pair(SA[i][j], SA[k][l]){
   couples = couples + pt(SA[i][j]) + pt(SA[k][l]);
   if ( NOT match(i, j, k, l) )
      sum = sum + pt(SA[i][j])+ pt(SA[k][l]);}
 return (sum / couples);}
```

Visual quality metrics usage

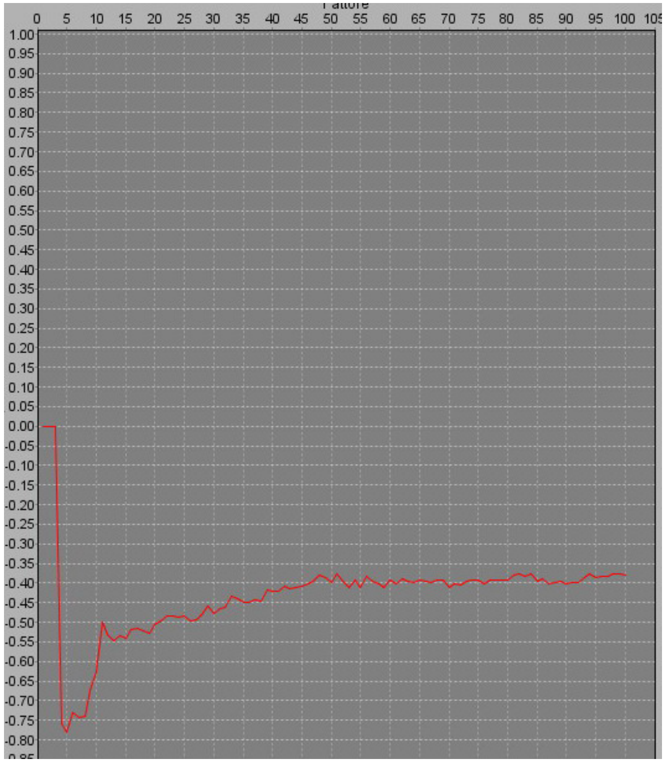
- Threshold comparison
  - to discard/accept the visualization
  - to activate ameliorating algorithms
- Comparison and evaluation of algorithms and visual techniques

- Algorithm driving

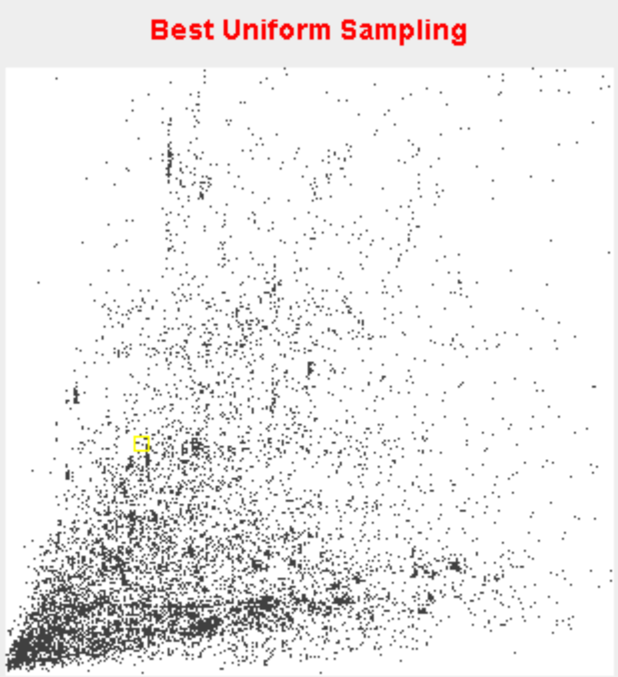
Practical usage of quality metrics



1-WPLDDr=0.392
i.e., 60% hidden



Best uniform sampling minimizing the metric



1-WPLDDr=0.738
i.e., 26% hidden

Conclusions

- Visual quality metrics **definition** and **usage** are critical issues
- A better comprehension of **what** we are measuring and **how** to use the measures is needed
- We believe that features preservation metrics produces **objective** indications about an image quality
  - but their definition is not a simple task