Text from Dave – email:

Ah, I thought you were talking about histograms.

For the scatter plots (which might also be called starfield plots), I don't think it's that zero values aren't important, but rather that the multiplicity of points of the form (x, 0) might not be of interest. The issue we ran into here in trying to quantify the "goodness" of a plot relative to the original is that this dataset is heavy with points of the form (x, 0) and (x, small number). So if we compare distributions between frames and original, we don't do well because we suppress points where there isn't much change in dye concentration, which happens a lot when dye concentration is zero.

But given that someone decided to do a scatter plot indicates that they aren't too concerned about multiple points with the same or close values, as you couldn't discern those on the plot. So I imagine the viewer is trying to learn something about the "spread" of the points -- which combinations of values are occurring. What are the depths where I'm seeing high concentrations? What's the deepest the dye is being detected. So, for such a viewer, a reasonable measure would be image similarity, which Michael's "grid measure" captures.

Also, I include text from past email, which you should look at if you didn't previously. It endorses a similar view of what a better approximation to a scatter plot is, though it has no better justification than intuition. But it's something we can site as support.

------------------------------------

From previous message

------------------------------------

See the text that goes with Figures 1-3 in this paper.

http://www.vldb.org/conf/1999/P15.pdf

It's an example of whether one point plot is better than another at approximating the plot of the full dataset. On the downside, they don't give any quantitative metric to say one is better than the other, just that one "generates a much better feel".

On the upside, if Yannis didn't know of an existing metric for this, then it's evidence that one doesn't exist. Also, it's something we cite to bolster providing a scatter plot as a task.

Also, need to check out where "starfield" leads.

There was another, simple case, where the task is just to approximate a signal -- be able to do a reasonable line plot. The metric would be something like maximum or RMS error.   
  
For episodes, there was a second metric, namely were episodes missed or falsely reported. We consider both "duration" and "existence" errors.   
  
For scatter plots, see my previous email. We are after the "spread" of points, and Michael's grid measure seems reasonable there. (Note that looking at spread is similar to looking at unique values in a single-valued dataset. For example, suppose I have a database of light bulbs from various manufacturers. I might be interested in looking at the distinct values in the "wattage" column, but not really be interested in how many different 40W bulbs there are.)\*   
  
For histograms, the metric was earth-mover distance.   
  
\*A long time ago, there used to be 37W light bulbs. Who can tell me why?