Eurographics / IEEE Symposium on Visualization 2011 (EuroVis 2011)
H. Hauser, H. Pfister, and J. J. van Wijk
(Guest Editors)

Volume 30 (2011), Number 3

Visual Boosting in Pixel-based Visualizations

Daniela Oelke[1], Halldor Janetzko[1], Svenja Simon[1], Klaus Neuhaus[2], and Daniel A. Keim[1]

[1] Data Analysis and Visualization Group, University of Konstanz, Germany
[2] Chair of Microbial Ecology, Technical University of Munich, Germany

Abstract

Pixel-based visualizations have become popular, because they are capable of displaying large amounts of data and at the same time provide many details. However, pixel-based visualizations are only effective if the data set is not sparse and the data distribution not random. Single pixels - no matter if they are in an empty area or in the middle of a large area of differently colored pixels - are perceptually difficult to discern and may therefore easily be missed. Furthermore, trends and interesting passages may be camouflaged in the sea of details.
In this paper we compare different approaches for visual boosting in pixel-based visualizations. Several boosting techniques such as halos, background coloring, distortion, and hatching are discussed and assessed with respect to their effectiveness in boosting single pixels, trends, and interesting passages. Application examples from three different domains (document analysis, genome analysis, and geospatial analysis) show the general applicability of the techniques and the derived guidelines.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques—Standards; I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms

1. Introduction

Nowadays, in many applications there is a need to analyze huge amounts of data. Among the popular techniques for visualizing large amounts of data are pixel-based approaches because they are able to display large data sets at a high resolution (see e.g., [Kei00, KSS07, LGP*07]). In the context of this paper, under 'pixel-based visualizations' we understand visualizations that use some small area to encode one data item. Note that the areas do not necessarily have to be display pixels but small rectangles. Pixel-based visualizations have been applied in many different domains including but not limited to geography [PSKN06], network and sensor data analysis [RG10, FN05], and document analysis [KO07]. Although very powerful, the technique comes with the disadvantage that interesting values or passages may easily be missed if the data is sparse. Besides, interesting details may be camouflaged in the sea of details.

Imagine you search for a term in a document and you are interested in the text passages which include this term. Typically you would use the find functionality of your favorite text viewer and the program would highlight the search terms. If we want to visualize the whole document with a pixel-based visualization (e.g., to fit it on the screen) and apply a term-based search, we have to somehow highlight the search results. We call this highlighting of pixels *boosting*. According to what we want to boost – single pixels (e.g., found search terms), passages (e.g., coding regions in genome data), or trends (e.g., the overall readability of sections while showing the complexity of each single word) – different boosting techniques perform best.

In this paper we discuss a number of approaches for visual boosting of pixels, passages, or an inherent trend. We distinguish between two definitions of importance. First, the importance of a data item can be defined by the numerical data value itself. In this case, the importance of a pixel is already encoded by the color-mapping, e.g., a single red pixels in an area of green pixels. Thus, we boost information that is already present in the pixel-based representation but not yet visually salient enough. In the following, we refer to such cases as **image-driven boosting**. Second, importance of a pixel can be defined in terms of external meta-information. In this case, the importance is not already represented by the color-mapping and needs to be encoded in the first place. An example is the boosting of a passage with highly divergent values or e.g., boosting of pixels that represent a noun - something that is only known by meta information. We refer to such cases as **data-driven boosting**. One characteristic of this type of boosting is that the values that have to be boosted usually span across the whole color scale - a fact which prevents using boosting techniques that are based on an adaption of the color scale.

This paper introduces and discusses different approaches for boosting pixels or passages. The techniques are reviewed with respect to their capability to amplify a single item, a passage, or a trend. Chapter 2 introduces different techniques for boosting and discusses their strengths and weaknesses. Next, in chapter 3 we describe the perceptual issues as the theoretical foundation for visual boosting. This is followed by an application section that exemplifies visual boosting by analyzing data from three different domains: document analysis, genome analysis, and geospatial data analysis. Chapter 5 classifies the different boosting techniques according to the application tasks they are most useful for.

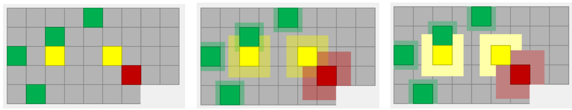
2. Boosting Techniques

Much research has been conducted in the domain of cognitive science to find out how eye movement works and what makes some objects stand out against others (see for example [War08]). It is assumed that we preprocess scenes by means of low-level properties such as color, shape, orientation etc. to decide where to direct our view. This can be exploited, for instance, in a search scenario in which the object has known properties that discriminate the item from the rest of the scene (e.g., when searching for a tomato, we search for a red object). The higher the feature-level contrast between the object and its surroundings, the easier it becomes to spot it.

In the following, we review some of the fundamental visual properties mentioned by Ware [War08] with respect to their applicability for boosting in pixel-based visualizations. Please note that many of the following techniques like distortion or hatching are highly dependent of the size of the pixel representation and cannot be applied to very small pixels.

2.1. Boosting with Halos

Size is an efficient and intuitive way to enhance the visibility of interesting or important pixels. However, in most cases we do not want to change the layout of the pixel-based visualization and thus have to overplot neighboring pixels. We therefore follow Colin Ware's suggestion of adding a surround color [War08]. To differentiate the halos from the neighboring pixels, it is best to use a less saturated or semi-transparent color. Using transparency comes with the advantage that even in overlap areas all halos are visible but may lead to blended colors that are difficult to interpret. Figure 1 exemplifies the different variants.



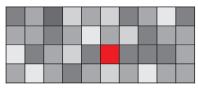
(a) (b) (c)

Figure 1: *Visual boosting with halos. Halos may be painted with semi-transparent or opaque colors. As the halos may overlap each other, the painting order is important. (a) Without visual boosting, (b) Boosting with semi-transparent halos, (c) Boosting with opaque halos.*

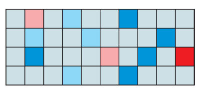
In most cases it is advantageous not to overpaint pixels representing a value in order to avoid camouflaging them. But still in areas in which several colored pixels are close to each other, their halos may overlap. In this case, the painting order is important. This is especially true when non-transparent halos are used but also for transparent ones, since the last color will be the most salient one.

For determining which pixels to boost, we can often measure a degree of importance. Boosting with halos has the advantage that the degree of boosting can be controlled. Changing the size of a halo has an immediate effect on the visibility of the boosted pixels.

2.2. Boosting with Colors



There are two possible ways to use color for boosting interesting or important pixels. The first way is to adapt the color scale in a way that important values are represented by a color with a high perceptual contrast compared to the less important data values. Thus, the boosted pixels are colored in a way that they are pre-attentively visible. We therefore change our colormap and represent values which should be boosted with a color that has a high perceptual contrast to the non-boosted values. Contrast colors can be found with the help of color wheels (for instance in Adobe Kuler [ADO10] or Color Scheme Designer [Sta10]) or by measuring their perceptual distance (e.g., in the CIE color space [CIE78]). The more heterogeneously colored the visualization is, the larger the perceptual distance of the boosting color has to be to ensure a good boosting effect (see [War08]). An easy way to achieve this is to use a black and white colormap and color important pixels with another color (e.g., red or yellow).

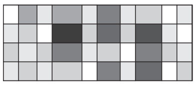


If the data set is sparse, not every pixel will be colored in the visualization. This means that we have empty pixels that do not encode a value. The second approach exploits this by coloring the empty pixels according to the value that should be boosted.

Typically, the boosted value represents the local or global trend (e.g., the average value) and is visualized by a less saturated color. As a consequence, pixels with colors similar to the background color become less salient. This is why it is advisable to use the technique only to represent a trend that is present in the data anyway.

2.3. Boosting with Distortion

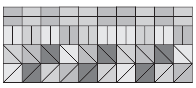
Distortion of pixels can be seen as another way of using size as visual property. By shrinking or enlarging pixels according to a user-defined importance measure, the focus is shifted to the larger ones while preserving the context, which is important for many application scenarios. This comes with the advantage of increasing the scalability because less space is used for unimportant passages.



We may also distort single rows or columns according to the number of boosted pixels or according to the sum of their importance values. This is only reasonable if a row or column is a meaningful and coherent unit. The distortion of columns, for example, would be reasonable if each column represents one point in time and each row shows a particular variable over time.

In other applications, such as geospatial analysis tasks, it may be favorable to distort the local neighborhood of the pixels to be boosted. This can help to decrease overlap that is for instance caused by enlarging important pixels in the geographical space. In section 4.3, we illustrate this for an example from the domain of geospatial analysis.

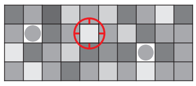
2.4. Boosting with Hatching



Hatching is a boosting technique which can only be used if the data values are mapped to small areas and not single pixels of the display. By using different hatching directions it is possible to distinguish several semantically connected groups of pixels. For instance, the two diagonal lines or vertical and horizontal lines may be used for hatching. Apparently the difference between vertical / horizontal lines is much more visible than the different diagonals, which may be a result from our culturally-inherited reading directions. Typically, the amount of hatching is also used to represent a quantitative value. However, this does not make sense for pixel-based visualizations where not many different degrees of hatching can be used due to the limited amount of space.

2.5. Boosting with Shapes or Glyphs

It is easy to spot one triangle in a set of rectangles or to spot the destination in a route planner when it is marked with a flag. But glyphs or shapes have one major problem: They occupy a different area than the original pixels. This results in different local contrasts and – even worse – in different color perceptions. It is therefore difficult to use colored pixel representations combined with different shapes while preserving the proper color perception. Furthermore,



both glyphs and the usage of a different shape (non-rectangular) requires that the pixel representation is large enough.

2.6. Other boosting techniques

In the previous sections, we introduced several techniques for boosting in pixel-based visualizations. Guided by the different visual properties presented in [War08], we selected the methods that seemed most applicable for enhancing the visual saliency of pixels, passages or trends. Nevertheless, the list is far from being complete. For instance, a user study that was conducted by Kosara et al. suggests that blurring of unimportant data points could be effective as well in a boosting task (see [KMH*02]). The study shows that sharp objects in a blurry background can be perceived preattentively.

Purposefully misaligning specific pixels might be effective as well but can only be used if there are gaps between the pixel rows. Other visual properties such as joined lines or curvature, which are mentioned in [War08] as well, are not usable for pixel-based visualizations.

Animations are one of the most eye-catching boosting techniques, as motion detection is hard-wired in our peripheral vision system. It is therefore possible to use, e.g., flashing of pixels to highlight important values. But animation has to be used very carefully as it can be very annoying if for instance many items on the screen are animated.

More interesting approaches for visual boosting using scatterplots are described by Pritzkau et al. in [PRSB10]. The paper describes the use of animated jittering in order to show coherent clusters in a pre-attentively manner. Additionally, the authors suggest to use illusionary motion to visually group data-points. This rotation effect is created by special shading techniques. For pixel-based visualizations, this technique is not well scalable (with respect to the size of the pixels), but could still be useful if only few pixels are to be displayed.

3. Perceptual issues

When using boosting techniques it is very important to consider the perceptual impact of changing the visual properties when boosting certain pixels. For example, when using background coloring, we have to be aware that the additional color will influence how the single colored pixels are perceived. In general, it can be said that it is best to use different visual channels (shape, color, motion) to let certain aspects pop out [War08]. If the same visual channel is used for boosting, a larger change in the feature space has to be made to make the object visually distinct. The same applies if the background is inhomogeneous, which is usually the case when working with pixel-based visualizations.

What does this mean for visual boosting? When coloring the background in sparsely populated visualizations, pixels with similar color will be more difficult to spot. At the same time, the larger the deviation of a pixel's color is from the background color, the more salient it becomes. In some cases this is exactly what we want to achieve because pixels that depict similar values compared to the background are already represented well enough. In all other cases, it is recommendable to use less-saturated colors for the background, which implies that we must use a color scale for coloring the pixels that is homogeneous with respect to the saturation.

Another boosting technique in which the perception of colors can interfere with the boosting are the glyphs. Since glyphs are often larger than the pixel itself, they may overlap with neighboring pixels. If the color of these pixels does not contrast with the glyph color, this decreases their effectiveness in terms of visual boosting.

Some of the boosting techniques work well for large pixels, but their effectiveness decreases for smaller scaling factors or the technique even becomes useless if single display pixels are used for each data value. For instance changing the shape of some of the pixels from rectangles to circles results in the fact that only part of the available area is filled. Otherwise put, part of the rectangular pixel area is filled with the background color. On small scale humans tend to mix these colors. Thus, red circles surrounded by a white background color will look brighter than rectangular pixels with the same color. The scalability of the different boosting techniques is further discussed in section 5.

Finally, literature on conjunctions of the visual properties must be taken into account. Ware [War08] states that for "rapid pop-out searching" not more than two different symbols should be used because otherwise spotting the differences will not be pre-attentive anymore. Since pixel-based visualizations already use color per default as one visual property, using several different boosting techniques at the same time to stress different aspects must be advised against. However, it can be very helpful to use a combination of different techniques to boost the *same* aspects, as this further increases the visual saliency of the pixels.

4. Application Examples from Different Domains

In this section we show application examples from three different domains, where boosting can help to increase the usefulness of a pixel-based visualization. Each application scenario exemplifies a different aspect discussed in the paper. The presented document analysis tasks show the challenges that a sparse data set comes with. Additionally, in the case of review analysis the inherent trend is important but not easily obvious. When analyzing genome sequences, the distinction between passages that encode a gene and those that have no known functionality is important. Genome analysis therefore is an example for enhancing the visibility of passages and for data-driven boosting. Finally, geospatial analysis has been chosen because of the restrictions that it poses on the placement of the pixels which increases the problem of overplotting and prevents using some of the techniques.

4.1. Document Analysis

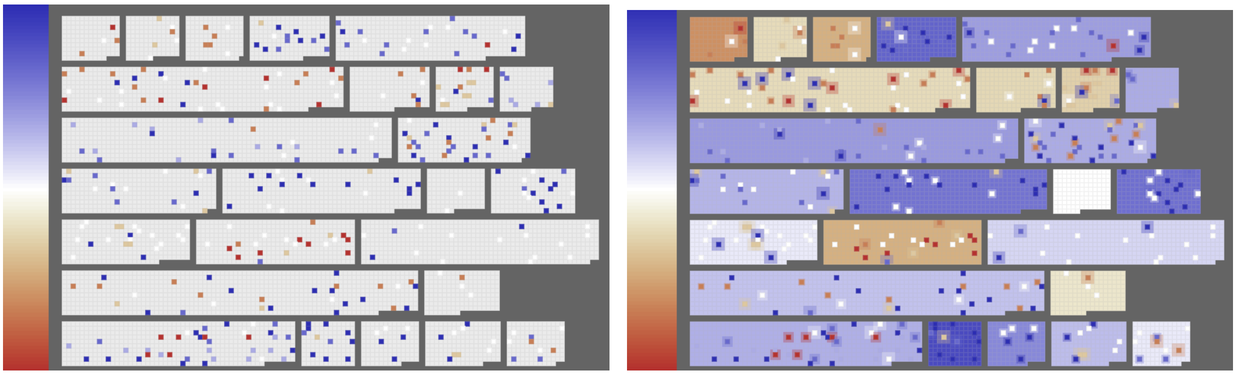
In the following, two examples from the domain of document analysis are given. The presented applications are typical representatives of pixel-based visualizations in the sense that the task does not require a special layout of the pixels (except for the sequential nature of the data). In the scenarios below, the sparseness of the data is the most challenging aspect in the visualization process.

4.1.1. Review Analysis

Customer feedback data is a valuable resource for both potential customers and companies. In contrast to overall ratings of a product, reviews come with the advantage that they reveal the strengths and weaknesses of a product in detail by commenting on certain product features. Various algorithms exist for identifying product features in natural language comments as well as to determine the sentiment that was expressed on them (e.g., [PL08, TM08, DL07, KH06, PE05]).

Figure 2 shows a visualization of customer feedback data. Each review is represented by a block of pixels where each pixel represents a word. Color is used to highlight negative (red) or positive (blue) mentions of product features. Different levels of brightness are used to encode the strength of the expressed opinion. Note that [OBK*08] uses a pixel-based visualization for review analysis, too. But in contrast to our scenario they display the data on review level instead of word level which results in a dense data set.

In figure 2(a) no boosting was applied. Although all the necessary information is present, it is difficult to identify trends and outliers in the data set. Figure 2(b) shows the same data with boosting by background coloring and halos. The background color is mapped to the average sentiment in the review, permitting to spot the trend of each document at a glance. Halos are used to make single colored pixels more salient. This allows the user to investigate if the trend is caused by a set of opposing opinions or by very similar ratings for all features. Using halos as a boosting technique permits to influence the boosting strength of each pixel separately. In this case, we decided to determine the halo sizes locally (for each review separately) and bias the visibility of the pixels towards outliers (rare colors). An alternative would have been to give the strongest boosting



(a) Without visual boosting (b) Local boosting with background coloring and semi-transparent halos

Figure 2: *Visual boosting of customer feedback data. Positively mentioned features are colored in blue, negatively mentioned ones in red. In (b) background coloring is used to emphasize the local trend and outliers are boosted with halos.*

weight to the pixels whose values deviate most from the trend.

4.1.2. Highlighting Search Terms

Part of a scientific paper is visualized in figure 3. Each pixel represents a word and words are grouped into sentences. In this case the task is to spot passages that are dense with respect to certain search terms. Pixels that represent the term "visual" are colored in purple, findings of the search term "human" in orange, and occurrences of "multimedia" are marked in green. Again the halo technique is used for boosting. But this time, halo sizes are determined globally. The rarer a search term is in the document, the higher is its boosting factor. Globally seen, even the most frequent term "visual" is still sparse. This is why we boost the occurrences of this term, too. The parameters of the halo technique permit to determine how much the boosting should be biased towards rare occurrences. In this case, the boosting factor of the rarest term is chosen to be about four times as large as the one for the most frequent search term.

4.2. Genome Analysis

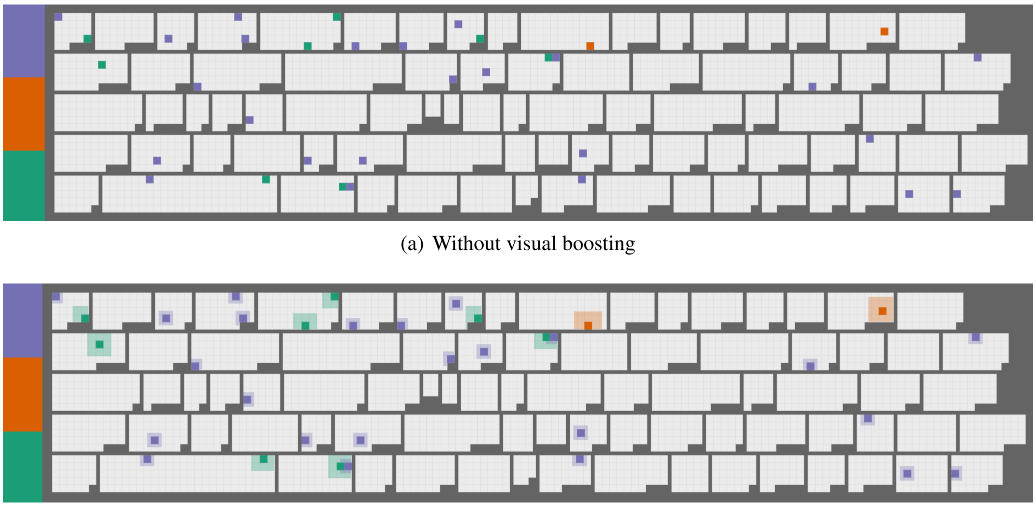
The genome of an organism is its hereditary information and represents (together with the cellular machinery) the 'construction plan'. The genome is a sequence over an alphabet of four letters (A, C, G, T - nucleotides). There are regions (genes) which code for known functional entities (proteins). However, for some sections it is not known if they code for a (new) protein or whether this section is indeed non-coding. To produce a protein, a gene containing segment of the genome is transcribed in a blueprint, called messenger RNA (mRNA). This mRNA is translated by the cellular machinery to the encoded protein.

With further development of sequencing technologies (next-generation sequencing technologies), it is now possible to sequence the whole transcriptome (all RNA) of a cell, RNA-Seq(uencing) [Met10, WGS09]. Since mRNAs are translated in proteins, it is possible to infer from this transcriptome, at least to some amount, the protein content of the cell, and to possibly identify new coding regions.

Thus, RNA-Seq is a promising new technique, but there are problems in data analysis. Due to technical reasons, only short segments of, e.g., 50 nucleotides of the RNAs can be sequenced using the SOLiD system [Met10]. The RNA, which can be a few thousand nucleotides in length, is therefore fragmented before sequencing. Each 50 nucleotide read, of which a few millions are produced in a sequencing run, has to be mapped to the genome. Sections of the genome that receive coverage from such reads are transcribed at this position. Because the reads from a certain long mRNA-molecule are obtained statistically, the coverage of a transcribed genomic region can be uneven or interrupted.

We use those RNA-Seq data as example of a task that requires data-driven boosting of passages. In figure 4, each pixel represents a nucleotide and the mapped color is dependent on the coverage. Thus, this reflects the relative amount of different RNA molecules that were found in the cell at the time of the analysis. The start and stop positions of known genes are given as meta-information. These gene stretches are marked with backslashes '\'.

Figure 4(a) shows the transcription of a segment of a genome. The coverage of the first gene depicted in the graphic is uneven and interrupted, but the transcription level is quite high compared to the background level. The surrounding regions are also transcribed which is common for most genes since those regions contain further information about where to start and stop protein synthesis of a gene on the mRNA. Interestingly, transcribed regions without annotated or suspected genes exist. Biologists are interested in finding



(a) Without visual boosting

(b) Global boosting with semi-transparent halos

Figure 3: *Visual boosting of search results. The visual saliency of sparse pixels is increased by halos.*

and describing those regions because they might contain new genes (or represent regulatory RNA). The last gene shown on the genome segment is apparently not transcribed.

Figure 4(b) shows the same segment using distortion to compress regions which are no genes and which are not transcribed. Furthermore, we use a bipolar colormap here to further emphasize the distinction between transcribed regions with and without annotated genes. Distortion is used to reduce to some extent the challenge of displaying long genomic sequences. Passages in which no transcription takes place are not interesting for the analysis. However, if we would remove them completely, the context information (gene neighborhood, length of the transcriptional gap, etc.), which is necessary for the analysis, would be missing. By shrinking these passages, the context is retained, but at the same time larger scalability and an emphasis of the important areas is achieved.

4.3. Geospatial Analysis

Boosting interesting or important data points in pixel-based visualizations of geospatial datasets is different from the previously described application examples. These datasets come with a geographical information, which inhibits the use of arbitrary layouts. Thus, geospatial datasets are not sequential but two-dimensional.

Some of the boosting techniques introduced in section 2 can easily be applied to geo visualizations such as halos, glyphs, or animation. Other techniques, such as hatching or changing the shape of pixels, might be less applicable because they require regular, closed areas of pixels to become dominant. In the following, we show an example of applying the halo boosting technique to a geographical data set and point out how the special challenges this comes with can be addressed.

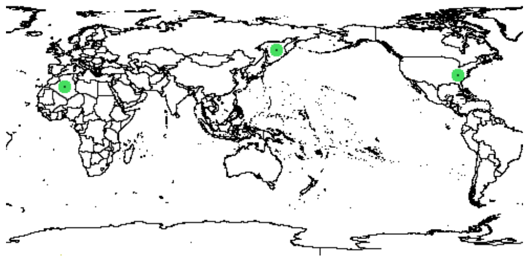
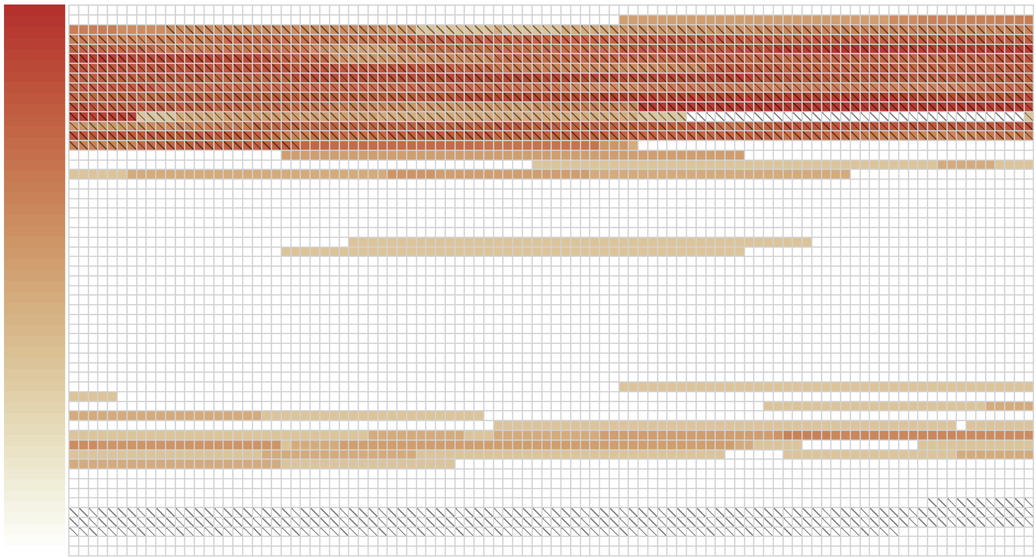


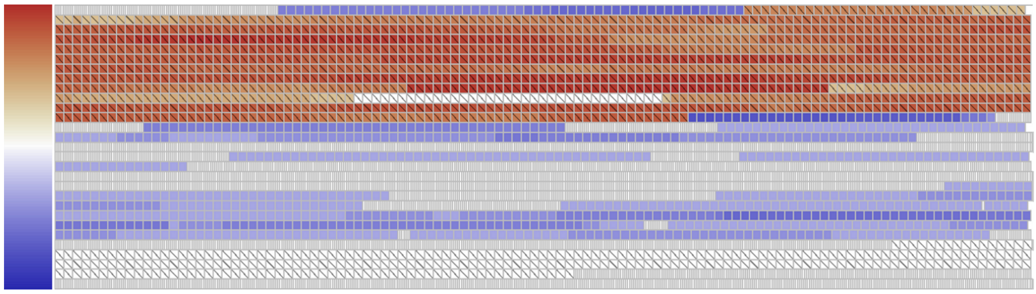
Figure 5: *Boosting of single points with circular halos. This approach induces overplotting, which has to be dealt with in more dense data sets.*

Figure 5 shows an example of boosting three single locations on a world map using circular halos. This works nicely because the data set is very sparse. However, in many application scenarios the data is not equally distributed across the earth but accumulates in certain regions. Imagine an earthquake data set that contains information about the location and strength of all occurring earthquakes in a certain time period. In regions of the world that are more likely to be struck by earthquakes, overplotting can easily occur in a pixel-based visualization. The problem is exacerbated if we use a boosting technique that needs space around the original pixel such as adding a halo or using glyphs.

One possibility to deal with the overplotting issues in a geographic dataset is to apply a local pixel replacement strategy. [KPSN03] and [KHD*09] present approaches for pi-



(a) Visual boosting of passages that represent gene stretches by hatching.



(b) Additionally, uninteresting passages that do no represents genes and are not transcribed are deemphasized by compression.

Figure 4: *Visual boosting of transcriptome data of Escherichia coli O157:H7 EDL933 using the SOLiD 4.0 technology. Shown are genes L7065 overlapping L7066, and L7071 from the plasmid pO157 (NC_007414). Bacteria were grown in M9 minimal medium.*

xel replacement. The key idea is to slightly shift pixels from their original position to some empty space if they overlap with some other pixel(s). Thereby, the geospatial topology is preserved as well as possible. Figure 6 was drawn using an adaptation of the methods that can deal with pixels of different sizes. Pixels with halos are placed first (as they are the largest ones and therefore most difficult to place). Pixels that would be overplotting if placed at their original position are then relocated to the nearest empty space.
Note that using semi-transparent halos would not help here because the local density of the points is too high. This inevitably would result in significant overlap between the halos with difficult to interpret blended paint.

Figure 6 shows a visualization of an earthquake data set, captured by the United States Geological Survey during a period from May 26th, 2010 to June 2nd, 2010 [USG10]. Halos are used to boost the visibility of the strongest earthquakes. The color scale shows the three different colors that are used to encode the magnitude of the earthquake together with their (less saturated) halo colors.

If the data set is even denser, relocation might not be possible anymore without enlarging the geographic region that the points are to be mapped to. The geo visualization community developed many local distortion algorithms that can be useful in such a situation. See [Tob04] for an overview of cartogram algorithms or [KPSN03] for an example that uses distortion algorithms to allow for better pixel placement.

5. Comparison of Boosting Techniques

In this section, we compare the different boosting methods of section 2 and present a guideline when to use which visual boosting method in pixel-based visualizations.

	image - driven boosting			data-driven boosting			trend		effectiveness in boosting	changes to layout	resolution dependency
	pixel		passage	pixel		passage					
	sparse	dense		sparse	dense		sparse	dense			
halos	+	-	-	+	-	-	o	-	+	o	+
background coloring	o	-	-	-		-	+	-	+	+	+
hatching	-		+	-		+	-		+	+	o
color map	+		+	-		-	-		+	+	o
animation	+		o	+		o	-		+	+	o
distortion	o		+	o		+	-		o	-	o
glyphs	+	o	-	+	o	-	-		o	+	-
shape	+		+	+		o	-		o	+	-

Figure 7: *Comparison of different boosting methods.* **+** *means the technique is well applicable for the specific task and* **o** *denotes medium effectiveness. Boosting techniques rated with* **-** *should not be applied to the respective application problem.*

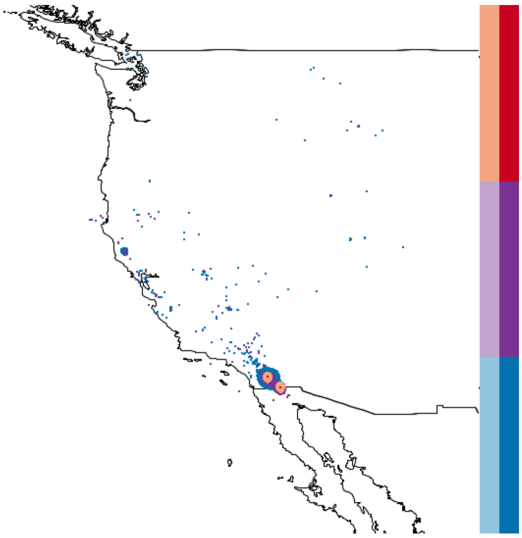


Figure 6: *Boosting of the strongest earthquakes with halos while reducing data- and halo-induced overlap. Color was mapped to magnitude, halos use less saturated colors.*

Boosting methods can be applied to increase the visual saliency of single pixels and passages or can be used to boost a trend. Furthermore, when assessing the effectiveness of a boosting technique, we also have to take into account if the data to be shown is sparse or dense. This is an important characteristic of the data because some techniques require empty space around a pixel. In addition, we distinguish image-driven and data-driven boosting as introduced in section 1. In addition to the general suitability of the boosting method under different conditions and for different use cases, we assess the methods regarding their resolution dependency and the strength of the boosting effect. We additionally consider the effects of the applied boosting technique to the layout, i.e. how much the layout has to be changed.

The best way to assess the effectiveness of the different techniques would be to conduct a large user study. Ideally, the different boosting tasks (image-driven, data-driven) and the type of data (sparse, dense) should be treated separately. Furthermore, combinations of boosting techniques should be taken into account as they may interfere with each other. Such a large study is clearly beyond the scope of this paper. Instead, we systematically collected the different aspects that have to be taken into account - thereby paving the way for further in-depth studies. Furthermore, we classified and assessed the techniques theoretically by taking the literature on perception into account. While for some aspects the decision seems to be clear (e.g., background coloring can only be used for comparably sparse data), other questions need to be answered by a future in-depth user study.

Figure 5 shows the resulting comparison table, part of which is discussed below.

5.1. Some additional comments

Although shapes can be considered as a special type of glyphs (see section 2.5), it is reasonable to make a distinction when evaluating their strengths and weaknesses. Glyphs cross the border of the boosted pixel while the different shapes (circles, triangles, etc. instead of rectangular pixels) stay inside the pixel boundaries. Hence, glyphs work best with empty surrounding pixels whereas shapes are not affected by the density of the data.

Another technique that is in need of special explanations is animation (or blinking). Animation certainly has a high effectiveness as a boosting technique. Nevertheless, we decided to downrate the value of animation because many people feel annoyed if too much of a scene is blinking. It is therefore important to use it with care. Furthermore, blinking comes with the disadvantage that it affects the color perception if the background color is visible from time to time. Another aspect that has to be taken into account is that animation can only be used in interactive applications and not for static media.

Finally, it has to be mentioned that distortion is a boosting technique that requires a change in the layout. This may decrease the users ability to follow the course of the values in an interactive application.

5.2. Sparse vs. dense data sets

The distinction between sparse and dense data sets is important for all techniques that use the space around the pixel for boosting. This is the case for halos, many glyphs, and of course background coloring. Halos and glyphs can be painted across neighboring colored pixels. However, this inevitably results in obscuring part of other pixels.
When boosting passages we do not have to make the distinction between sparse and dense because the considered segment of the data set is per definition not sparse.

5.3. Image- vs. data-driven boosting

Image-driven as opposed to data-driven boosting visually strengthens information that is already present in the original plot. It is basically a redundant encoding for part of the pixels which is the reason why adapting the color map does work for image-driven boosting but not for data-driven boosting. In the latter case, every adaption of the color map would inevitably destroy the information that was mapped to color or at least significantly impede reading the visualization. Similarly, using background coloring in a data-driven scenario would mean to use the visual property *color* twice, once for the values themselves and second for a summary measure or to encode some meta-data.

For the visual property *shape*, the difference between image- and data-driven boosting has another reason. Changing the shape usually results in empty space around a pixel. Recall that we inscribe circles or triangles into the rectangular space reserved for the pixel. Thus, especially for small scale representations the color perception is affected because we tend to mix the colors of small neighboring areas with each other. Since in image-driven boosting all pixels with a specific color are changed, the distortion of the color perception is not as severe as it is for data-driven boosting.

5.4. Boosting single pixels vs. boosting passages

Some methods work well for boosting single pixels but not for boosting passages. The reason is that even sparse data is not sparse in terms of boosting if it contains continuous sequences of pixels. This impacts all methods that are in need of empty neighboring pixels. On the other hand, the perception of some methods depends on the rate of occurrence. One hatched pixel is significantly less salient than a whole sequence of hatched pixels. The same is true for distortion which benefits from the visual accumulation of a whole sequence of compressed or enlarged pixels. In contrast, most glyphs are visually very dominant. Besides, they usually cross the border of a pixel which means that displaying them in a sequence can easily result in a cluttered representation. If glyphs are applied for boosting passages, special glyph symbols for this task have to be developed.

5.5. Boosting of trends

There are several methods to boost pixels or passages, but there is only one good method to boost trends in data sets: the background coloring technique. Partly, also halos can be used for visualizing trends. If the most frequent values instead of rare values are enhanced with halos, this results in an effect similar to background coloring. Note that this can only be done if the trend can be represented with the existing pixels and does not require additional calculations such as averaging the values. Since both techniques work best for sparse data, none of the introduced techniques can be recommended for displaying an inherent trend of a dense data set.

6. Conclusion

In this paper we addressed the problem of visually boosting interesting or important data points in pixel-based visualizations. Several boosting techniques such as halos, coloring, distortion, and hatching were systematically reviewed and compared to each other. Furthermore, we discussed perceptual issues that come with applying the presented techniques. Another contribution of the paper is the evaluation of the different boosting techniques with respect to their applicability in different tasks. We could show that tasks should be classified as working on sparse or dense data sets and with respect to the boosting task (boosting of pixel, passages, or a trend). Furthermore, a distinction between image-driven and data-driven boosting is important where image-driven refers to boosting that visually enhances information that is already present in the plot (redundant encoding) whereas data-driven boosting relies on importance measures that determine important pixels based on some meta-information or function. Several application examples were presented to show the wide applicability of the techniques and to exemplify the derived guidelines.

Acknowledgement

This work has been partly funded by the German Research Society (DFG) under the grant GK-1042, Explorative Analysis and Visualization of Large Information Spaces, Konstanz and the SPP 1335, sub-project Spatiotemporal Analysis of Movements as well as SPP 1395, sub-project Finding Overlapping Genes. Thanks to Richard Landstorfer (TU Munich) for providing the transcriptome data.

References

[ADO10] Adobe kuler. http://kuler.adobe.com/, 2010. Adobe Systems Incorporated. 2

[CIE78] C.I.E. Recommendations on uniform colour spaces, colour difference equations, psychometric colour terms. *Supplement No.2 to CIE publication No.15 (E.-1.3.1) 1971/(TC-1.3.)* (1978). 2

[DL07] Ding X., Liu B.: The Utility of Linguistic Rules in Opinion Mining. In *Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR)* (2007), ACM, pp. 811–812. 4

[FN05] Fink G. A., North C.: Root Polar Layout of Internet Address Data for Security Administration. In *Proceedings of the IEEE Workshop on Visualization for Computer Security (VizSEC)* (2005), IEEE Computer Society, pp. 55–64. 1

[Kei00] Keim D. A.: Designing Pixel-oriented Visualization Techniques: Theory and Applications. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 6*, 1 (2000), 59–78. 1

[KH06] Kim S.-M., Hovy E.: Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text* (2006), ACL, pp. 1–8. 4

[KHD*09] Keim D., Hao M., Dayal U., Janetzko H., Bak P.: Generalized Scatter Plots. In *Information Visualization Journal (IVS 2009)* (2009), Macmillan Publishers Ltd. 6

[KMH*02] Kosara R., Miksch S., Hauser H., Schrammel J., Giller V., Tscheligi M.: Useful properties of Semantic Depth of Field for Better F+C Visualization. In *Proceedings of the Symposium on Data Visualisation 2002* (2002), VISSYM '02, Eurographics Association, pp. 205–210. 3

[KO07] Keim D. A., Oelke D.: Literature Fingerprinting: A New Method for Visual Literary Analysis. In *Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology (VAST '07)* (2007), IEEE Computer Society, pp. 115–122. 1

[KPSN03] Keim D., Panse C., Sips M., North S.: PixelMaps: A New Visual Data Mining Approach for Analyzing Large Spatial Data Sets. In *IEEE International Conference on Data Mining (ICDM)* (2003), IEEE Computer Society, pp. 565–568. 6, 7

[KSS07] Keim D. A., Schneidewind J., Sips M.: *Scalable Pixel Based Visual Data Exploration*. Springer, 2007, pp. 12–14. 1

[LGP*07] Lévy P., Grand B., Poulet F., Soto M., Darago L., Toubiana L., Vibert J.: *Pixelization Paradigm: First Visual Information Expert Workshop, VIEW 2006, Paris, France, April 24-25, 2006: Revised Selected Papers*. Springer, 2007. 1

[Met10] Metzker M. L.: Sequencing technologies - the next generation. *Nature Reviews Genetics 11*, 1 (2010), 31–46. 5

[OBK*08] Oelke D., Bak P., Keim D. A., Last M., Danon G.: Visual Evaluation of Text Features for Document Summarization and Analysis. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST 2008)* (2008). 4

[PE05] Popescu A.-M., Etzioni O.: Extracting Product Features and Opinions from Reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (2005), ACL, pp. 339–346. 4

[PL08] Pang B., Lee L.: Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval 2*, 1-2 (2008), 1–135. 4

[PRSB10] Pritzkau A., Radloff A., Schumann H., Bartz D.: Scattering and Jittering : Using Real and Illusionary Motion for Better Visual Scatterplot Analysis. *Presented at the Poster Session at IEEE Information Visualization Conference 2010 (InfoVis), Salt Lake City, USA*, 2010. 3

[PSKN06] Panse C., Sips M., Keim D. A., North S. C.: Visualization of Geo-spatial Point Sets via Global Shape Transformation and Local Pixel Placement. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 12*, 5 (2006), 749–756. 1

[RG10] Rodrigues P., Gama J.: A Simple Dense Pixel Visualization for Mobile Sensor Data Mining. In *Knowledge Discovery from Sensor Data*, Gaber M., Vatsavai R., Omitaomu O., Gama J., Chawla N., Ganguly A., (Eds.), vol. 5840 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2010, pp. 175–189. 1

[Sta10] Stanicek P.: Color Scheme Designer 3. http://colorschemedesigner.com/, 2010. 2

[TM08] Titov I., McDonald R.: A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *Proceedings of ACL-08: HLT* (2008), Association for Computational Linguistics, pp. 308–316. 4

[Tob04] Tobler W.: Thirty Five years of Computer Cartograms. *Annals of the Association of American Geographers 94*, 1 (2004), 58–73. 7

[USG10] U.S. Geological Survey, Latest Earthquakes: Feeds & Data. http://earthquake.usgs.gov/earthquakes/catalogs/eqs7day-M1.txt, 2010. 7

[War08] Ware C.: *Visual Thinking for Design*. Morgan Kaufmann, 2008. 2, 3, 4

[WGS09] Wang Z., Gerstein M., Snyder M.: RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics 10*, 1 (2009), 57–63. 5