

An Evaluation of Visualization Techniques to Illustrate Statistical Deformation Models

J. J. Caban¹, P. Rheingans² and T. Yoo¹

¹National Library of Medicine, NIH, USA

²University of Maryland, Baltimore County (UMBC), USA

Abstract

As collections of 2D/3D images continue to grow, interest in effective ways to visualize and explore the statistical morphological properties of a group of images has surged. Recently, deformation models have emerged as simple methods to capture the variability and statistical properties of a collection of images. Such models have proven to be effective in tasks such as image classification, generation, registration, segmentation, and analysis of modes of variation. A crucial element missing from most statistical models has been an effective way to summarize and visualize the statistical morphological properties of a group of images. This paper evaluates different visualization techniques that can be extended and used to illustrate the information captured by such statistical models. First, four illustration techniques are described as methods to summarize the statistical morphological properties as captured by deformation models. Second, results of a user study conducted to compare the effectiveness of each visualization technique are presented. After comparing the performance of 40 subjects, we found that statistical annotation techniques present significant benefits when analyzing the structural properties of a group of images.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—Visualization and Uncertainty Visualization

1. Introduction

The large amount of anatomical variability often observed among healthy human subjects has created great interest in the analysis and understanding of morphological properties. Over the years, statistical models have proven to be effective methods to capture the properties of groups of images delineated by common characteristics. However, given the large inter-image variability often found within input data, as well as the complexity of most non-linear deformations, statistical models are often limited at enabling accurate exploration, analysis, and comparison of morphological deformation properties.

Those limitations have underscored the need for visual representations that can effectively illustrate the structural and morphological statistical properties of groups of images. This paper explores different visualization techniques to effectively show morphological properties as captured by statistical deformation models (SDMs). Recently, SDMs have emerged as a promising approach to capture the overall morphological information of a collection of 2D/3D images. SDMs use deformation fields – such as those obtained from

non-linear registration techniques – to capture the structural characteristics of a group of images.

In medical imaging, Rueckert et al. presented the original method for generating SDMs and demonstrated how they can be used to estimate the average model of a particular class [RFS03]. Albrecht et al. presented a technique to generate SDMs with minimum bias towards the reference image and demonstrated how such models can be used to enhance model-based registration [ALV08]. Over the last several years, many other techniques have been proposed and used to capture morphological properties including deformation- and tensor-based morphometry models, relational deformation models, and many others [CWP*01, LKJ*06]. In general, all of those deformation models can be described as variations of SDMs. Despite the recent popularity of SDMs, effective ways to explore, visualize, and summarize the large amount of structural information contained within such models have not been studied. Analyzing and visualizing the average model does not provide enough information to reach definite conclusions about the deformation properties of a particular group.

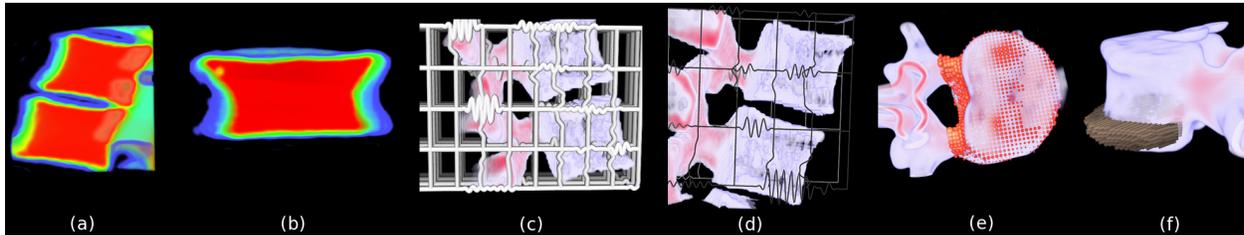


Figure 1: Results of using four illustration techniques to show the statistical morphological properties of an SDM model generated from the lumbar vertebrae of 20 subjects. (a-b) Results of using likelihood volumes to illustrate the probabilistic properties of a group of images. (c-d) Results of using deformable grids to show statistical deformation properties and characterize regions with high variability. (e) Results of using spherical glyphs to annotate the variability of different areas of the vertebra. (f) Results of using line-based glyphs to illustrate deformation range and morphological variability.

Over the last two decades, a number of annotation and illustration techniques have been proposed in the fields of scientific and information visualization [PWL97, WPL96, PCS05]. Some of those existing techniques can be extended to work with SDMs and function as methods to summarize structural deformation properties. This paper evaluates four visualization techniques that have been extended to meet the requirements and needs of statistical morphological analysis. First, *likelihood volumes* are introduced as a method to display the probabilistic properties of a group. Second, three existing uncertainty visualization techniques – *deformable grids*, *spherical glyphs*, and *line glyphs* – are extended to work with SDMs to show the structural variability as captured by the model. Finally, a user study is conducted to measure and compare the effectiveness of each visualization technique.

2. Previous Work

In scientific visualization, a large number of illustration techniques have been proposed to convey uncertainty information within 2D and 3D images. Some of those techniques are uncertainty glyphs, procedural annotations, pseudo-colorings, box plots, summary plots, probabilistic animations, and pointillism-based techniques [PWL97, WPL96, PCS05, PKR06, GR04]. Note that most techniques have primarily focused on illustrating the uncertainty that might occur within a single image.

Saunders et al. presented a method to visualize the statistical properties of high-quality simulation data by simultaneously visualizing the mean diameter and the standard deviation of nanoparticles within a single image [PCS05]. Their method used pointillism, circular glyphs, and color variations to convey statistical information. Wittenbrink et al. introduced *uncertainty glyphs* to visualize uncertainty in winds and ocean currents [WPL96, PWL97]. They were able to illustrate multiple metrics simultaneously including the direction, magnitude, mean direction, and length of the currents. However, their technique only worked with a single vector field. Cedilnik and Rheingans introduced procedu-

ral annotations including grids to show uncertainty information [CR00]. In addition, Grigoryan and Rheingans presented a point-based technique to visualize surfaces along with uncertainties [GR04]. The general idea was to displace points according to their uncertainty values. Potter et al. introduced a summary plot which overcomes some of the limitations found with box plots [PKR06]. Kindlmann proposed a new set of glyphs – superquadric glyphs – to overcome some of the asymmetry and ambiguity problems of regular glyphs [Kin04].

Uncertainty visualization techniques have also been used for direct volume rendering. Kniss et al. presented a visualization technique which takes into account uncertainty within the underlying data [KUS*05]. Ljung and Persson introduced probabilistic animations as a technique to display the uncertainty, error, and misclassification often introduced by transfer functions [LP07].

A key component of any visualization technique is to compare and measure its effectiveness at achieving what it was designed to do [Nor06]. Given the need for thorough and methodologically sound methods to compare visualization techniques, user studies have been accepted as suitable approaches to measure the effectiveness of different illustration techniques. During the last several years many research projects have tried to measure the potential benefits and limitations of different visualization techniques [KHI*03, TM04, HIL*05, Nor06]. For example, North and Shneiderman analyzed how multiple coordinated visualizations can enable users to rapidly explore complex information [NS00]. Laidlaw et al. presented a user study that compared the effectiveness of six visualization methods to illustrate two-dimensional vector fields [LKJ*05]. Bair and House presented the results of two controlled studies comparing the effectiveness of layered surface visualizations under various texture conditions [BH07].

In our effort to identify visualization techniques that can illustrate the information captured by statistical models we introduce *likelihood volumes* for SDMs as a technique to show the probabilistic aspect of different regions of the

model. In addition, we extend *deformable grids*, *spherical glyphs*, and *line glyphs* to work with SDM models.

3. Statistical Deformation Models

Given a set $\mathbb{I} = \{I_1, I_2, I_3, \dots, I_n\}$ with n linearly registered images, an SDM is built by performing a statistical analysis of the deformation fields required to map each input image $I_i \in \mathbb{I}$ to a common coordinate system [RFS03]. In medical imaging, the morphological differences between individual subjects or the structural differences caused by the progression of a particular disease cannot be estimated by a simple linear transformation. Often, the estimation of individual local transformations for every voxel $I(x, y, z)$ is required to align two images. Given an image I_i and reference image I_R , a non-linear image registration technique estimates a displacement vector \vec{v}_{xyz} for each voxel. The collection of displacement vectors \vec{v}_{xyz} forms a deformation field D' with a dense feature correspondence between the two images under consideration. From the registration results, the inverse field D can also be estimated with the displacement information required to map image features from the reference image I_R to the input image I_i . By registering each image $I_i \in \mathbb{I}$ with the common coordinate system defined by I_R , a set $\mathbb{D} = \{D_1, D_2, D_3, \dots, D_n\}$ can be estimated. The resulting set \mathbb{D} contains n displacement vectors \vec{v}_{xyz} for every voxel $I_R(x, y, z)$. That group of vectors can be used to extract the statistical properties and determine how a particular region can deform. By treating each point as an independent random variable, a multi-variate distribution $\mathcal{N}(\mu, \Sigma)$ can be estimated to create a probability density function (PDF) representing the local variability of the data. Since non-linear registration techniques cannot guarantee a one-to-one correspondence between the two images, *super-voxels* or *patches* can be used to more effectively estimate morphological changes of a particular region. In particular, an image can be divided into n patches $\mathbb{P} = \{p_1, p_2, \dots, p_n\}$, where p_i is a patch grouping k^3 voxels and $\mathcal{N}(\mu, \Sigma)_{p_i}$ a PDF with the statistical properties of that region.

4. Approach

Once an SDM is generated, the model can be described as an image consisting of a large collection of PDFs defining the morphological properties of each region $I(x, y, z)$ or patch $p_i \in \mathbb{P}$. SDM models have been used to generate the average model of a particular class, enhance image segmentation, and improve model-based registration [RFS03, ALV08]. However, users are interested in analyzing the information contained within the deformation models to better understand the characteristics of a group, which structures are related, and which regions present similar patterns. Such analysis can be done by creating effective visualization and illustration techniques capable of summarizing the statistical deformation properties of a group of images into a single picture. This section presents four techniques that can be used to illustrate the information captured by SDMs.

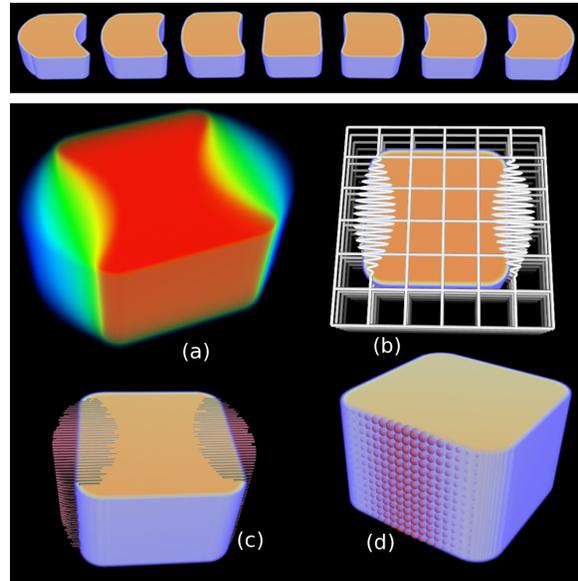


Figure 2: (top) Sample of one of the synthetic datasets used to test the different illustration techniques. Note that since the images are in order, we can easily see the deformation pattern of this collection. (bottom) Results of using four illustration techniques to show the statistical properties of our SDM model which was generated from 42 different 3D images. Results of using (a) Likelihood Volumes, (b) Deformation Grids, (c) Line Glyphs, and (d) Spherical Glyphs.

4.1. Likelihood Volume for SDMs

SDMs are effective at capturing the overall changes, variability, and characteristic aspects of the class under consideration. Directly from an SDM model, we can generate a probabilistic image which can be used to illustrate the likelihood of different structures and highlight areas that are more stable within the class under consideration. Such visualization techniques have significant applications in image registration, segmentation, and feature matching given that the stable areas of an image can be used to specify locations from where image features can be extracted and used to enhance image analysis.

A likelihood volume is generated as follows: First, an empty 3D image I_{LV} of equal or greater size as the global coordinate system I_R is created and initialized to zero. Then, by looping through each voxel or patch $p_i \in \mathbb{P}$ of the SDM under consideration, each PDF $\mathcal{N}(\mu, \Sigma)_{p_i}$ is evaluated to obtain the deformation range of the area under consideration. In particular, since each normal distribution \mathcal{N}_{p_i} includes the expected deformation vector $\vec{v} = (\mu_x, \mu_y, \mu_z)$ and its variance $\hat{\Sigma} = (\sigma_x, \sigma_y, \sigma_z)$, we can estimate all possible spatial locations (voxels) that the current region p_i can deform into. During the evaluation process, the image I_{LV} is then used to accumulate the number of visits each spatial location $I(x, y, z)$ receives. Finally, the image I_{LV} can be illustrated using regular rendering techniques to show the probabilistic properties

of a group of images. In particular, likelihood volumes can be used to illustrate the deformations that are more likely to occur and areas that have the highest probability of being included within any new image belonging to the same class.

Figures 1(a-b) show the results of visualizing an SDM of the vertebra using *likelihood volumes*. The SDM was trained using 20 CT images of the spine. The results show how *likelihood volumes* can be used to analyze the probabilistic properties of this particular anatomy. In addition, the images show how specific parts of the vertebra can vary among different subjects. Figure 2a shows the results of using *likelihood volumes* to illustrate the statistical properties of an SDM trained on 42 volumes of deforming cubes. Note that *likelihood volumes* are effective at showing areas shared by most cubes as well as illustrating the extent of the deformation fields.

Likelihood volumes have several advantages over other visualization methods. First, the complete set of PDFs can be used to estimate the likelihood volume. Second, the volume can be automatically extracted from an SDM without the necessity of analyzing any of the images used to generate the model. In other words, we can generate a volume directly from the abstract representation of an SDM which shows the stable and most variable regions of the group.

There are a few disadvantages of likelihood volumes. First, it can be challenging to perform simultaneous visualization of the likelihood model and a member volume of the SDM. Such a limitation can be solved by using a two-level volume rendering approach [HMBG01]. Second, it can be hard to estimate the average deformation directly from a likelihood volume. Third, the actual shape of the structure under consideration can be lost.

4.2. Deformation Grids for SDMs

To enable exploration, analysis, and comparison between an SDM and a given volume, it is important to create tools that can preserve shapes while illustrating statistical properties. An effective visualization technique should also permit simultaneous illustration of a volume while annotating it with information obtained from the SDM. The annotations, however, must not overshadow the underlying model or introduce misleading information. Also, effective annotation techniques must provide insights about the variability of the class under consideration.

We demonstrate how statistical deformation information can be conveyed using *deformation grids*. Previously, deformation grids have been used in image registration to show the specific transformation required to map features from one anatomical model to another subject. Similarly, in uncertainty visualization, deformation grids have been described as procedural annotations that can be used to show data variability [CR00]. We have extended deformable grids to work with statistical deformation models.

The general idea of deformation grids for SDMs is to overlay a grid pattern over the volume data and distort the lines in such a way that the underlying statistical deformations captured by the SDM can be illustrated. Areas with high deformation will have a wavy shape representing the range of the statistical deformation while areas with small deformations will be almost straight.

A deformation grid for SDMs is generated as follows. First, given a parameter d with the distance between two grid lines, a 2D or 3D grid is generated. When annotating the statistical properties with 3D grids, an opacity-based technique is used to reduce the intensity of the grid along the z direction. Such intensity modulation is done to avoid visual clutter which can cause confusion. Second, a set of parameters $s = (s_x, s_y, s_z)$ is specified to describe the sampling along each axis. At this step, every grid point that is a multiple of s_x or s_y or s_z is distorted to convey the statistical information as provided by the deformation model. For instance, to distort vertical lines, the PDF of each voxel that is a multiple of s_y is evaluated. Then, every even point is distorted to show the negative deformation range while every odd location is distorted to show the positive deformation. Note that the same process is done for each axis which enables the creation of deformation grids to analyze SDM models. The actual parameters μ and σ of each PDF become parameters that control the distortion and appearance of the annotation grids.

The biggest advantages of this method are that it can be easily computed, does not require complex filtering of the SDM model, and it can quickly illustrate the general statistical deformation described by a model. However, there are also a few disadvantages. First, deformation grids for SDMs only show the statistical information within the grid lines. Second, if the parameter d with the distance between two grid lines is smaller than the variability of the data, deformation grids can become dense and hard to understand.

Figures 1(c-d) show the results of visualizing an SDM of the vertebra using *deformation grids*. We can see that there are specific regions within the vertebra that can deform significantly. Also note that different line properties can be used to better fit the user's preferences and requirements. Figure 2b shows the results of using *deformation grids* to illustrate the statistical deformation properties of a collection of 42 cubes. Figure 3 shows two illustrations obtained by using *deformation grids* to visualize the statistical morphological properties of a longitudinal dataset of the brain and a collection of femur heads. From the results we can see that *deformation grids* can be successfully applied to 2D or 3D images.

4.3. Line-based Glyphs

Glyphs have been widely used in scientific visualization to show uncertainty of data [PWL97]. We have picked two glyph-based methods to show statistical deformation of a volume: *line-glyphs* to show deformation range and

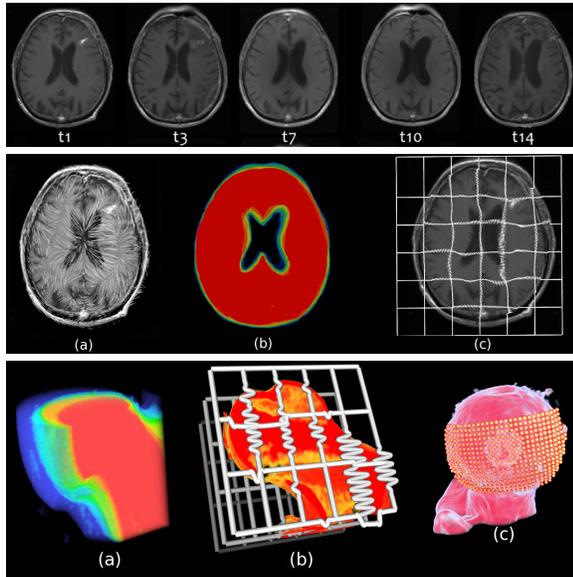


Figure 3: (top) A set of longitudinal images of a patient with low-grade gliomas which are often analyzed by physicians to measure deformation properties. (center) Sample annotations that can be used to summarize and analyze the statistical deformation properties of a group. (bottom) [a-c] Sample of the femur dataset also used within the survey.

spherical-glyphs to show data variability. Those glyph-based visualization techniques have been chosen because they are intuitive visualization methods, do not distract the viewer from the overall understanding of the data, and can effectively show how a voxel or a specific region behaves.

Given the set of voxels under consideration, the PDF $\mathcal{N}(\mu, \sigma^2)_{p_i}$ describing each spatial location p_i can be evaluated. By using the average transformation $\vec{\mu} = (\mu_x, \mu_y, \mu_z)$ and the covariance matrix $\hat{\Sigma} = (\sigma_x, \sigma_y, \sigma_z)$ the range of the deformation can be estimated. The range for a region p_i can be described as $R(p_i) = \mu_x \pm \sigma_x, \mu_y \pm \sigma_y, \mu_z \pm \sigma_z$. That range represents how the super-voxel under consideration behaves and the possible locations within which that voxel or region can be found in similar images. By evaluating each PDF and estimating its range, a line glyph can be generated for each voxel or region of interest, thus illustrating the deformation properties of the model under consideration.

We found that *line glyphs* can be used to illustrate the range of the deformations as well as to analyze the relationship between different objects. In medical imaging such visualization has the potential for allowing physicians to explore the data and compare normal and abnormal models. Given the large amount of deformation a voxel can undergo, we found that glyph-based annotations are more effective when a subset of the PDFs are considered by specifying an ROI, intensity range, or boundaries.

Figure 1(f) shows the results of illustrating the deformation range of a region of a vertebra using *line glyphs*. From the annotation we can analyze the range of the morphological deformations that can occur within the lower part of the lumbar vertebrae (L1). Figure 2c shows the results of using *line glyphs* to illustrate synthetic data. In this example, the lines show the range of the deformation along the boundaries of the cube.

4.4. Sphere-based Glyphs

Frequently, users are interested in knowing the overall variability of a particular region instead of knowing detailed information about the specific range of that region. We have extended spherical glyphs to work with statistical deformation models to convey information about the statistical variance. By evaluating the corresponding PDF $\mathcal{N}(\mu, \sigma^2)_{p_i}$ for the region p_i under consideration, we can estimate the range $R(p_i) = \mu_x \pm \sigma_x, \mu_y \pm \sigma_y, \mu_z \pm \sigma_z$ and by computing the magnitude of such a vector, we can get the variance of each voxel. Then, by mapping the variance to the size of a glyph, it is possible to illustrate the overall variance of a volume without introducing too many artifacts or information to the visualization system.

Figure 1(e) shows the results of visualizing an SDM of the vertebra using *spherical glyphs*. This annotation technique can be used to show variability and the likelihood of observing structural deformations within a particular region. In this illustration, we can see that the pedicles of the vertebral arch (bridges between the posterior and anterior structures) have a higher variability among normal subjects than the body of the L1 vertebrae. Figure 2d shows the results of using *spherical glyphs* to illustrate synthetic data. Figure 3(bottom)[c] shows the regions of the femur that are more likely to change among human subjects.

5. User Study

Visualization and illustration techniques have a direct impact on the user's ability to understand data and statistical models. We have conducted a user study to examine the effectiveness of our statistical illustration techniques and to obtain quantitative measurements to compare the different annotation techniques. The primary goal was to test our hypothesis that: *statistical illustration techniques improve the accuracy and confidence of user judgments at analyzing statistical morphological deformations*.

The user study consisted of three main sections. First, the *Introduction* where detailed information and explanation about each individual annotation technique was given. The main purpose of this section was to teach subjects about each of the visualization techniques. Second, a *User Preferences* section where 20 illustrations were shown and users were asked to score each annotation. In each question, a collection of images were shown together with the four corresponding

visualization techniques illustrating the statistical deformation properties of the group under consideration. Users were then asked to score each of the annotations techniques using a Likert scale with five choices ranging from poor to excellent. Figure 2 shows one of the synthetic examples used to capture the user preferences. See the supplemental material included with this paper for sample questions.

The third section was the *User Performance* section designed to compare how well subjects can infer the deformation properties of a group of images based on either the raw data or a single annotation. To capture and compare the effectiveness of each technique, 15 images representing different SDMs of synthetic and medical datasets were presented and users had to determine the morphological properties of the group under consideration directly from a single visualization. Annotations such as the illustrations in Figures 2(bottom) and 3(bottom) were presented and users had to answer a question about the statistical deformation properties of the group under consideration. For instance, from a single illustration from Figure 2(bottom) users had to understand that the input cubes were only deforming left and right and the center of the cubes were deforming more than any other region. In addition, to test how accurately users can infer the deformation properties of a group from raw images, groups of images – like the longitudinal dataset shown in Figure 3(top) – were shown and users had to pick the anatomical region that presented the largest morphological deformation. Figure 3 shows some of the real-world images used within this section.

To further compare the differences between visualization techniques, the time required to answer each question and the user’s confidence in the solution were recorded. All the questions within each section were displayed in random order to avoid any learning effect. In addition, a mix of synthetic and real-world medical 2D/3D images was used throughout the user study. For more information about the structure of our user study, the design, the datasets, and sample questions please see Section A of the supplemental material.

To complete our user study, 65 individuals from the department of Computer Science and the Medical Center were randomly selected and invited electronically to participate in our user study. The online survey was accessed by 51 unique subjects. Eleven subjects did not complete the entire survey and their partial answers were not included in our analysis. In total, forty participants completed the survey, thirty computer scientists and ten non-engineers with expertise in medical research.

6. Results

To measure the effectiveness of each illustration technique, compare their efficiency, and quantify the effects of a given approach with respect to other annotation techniques, the survey was analyzed using multiple statistical techniques.

First, each variable was tested for homogeneity by using Levene’s test [Lev60]. Then, analysis of variance (ANOVA) was used to better capture the differences among the four visualization techniques and obtain significance values.

All our tests started with our null hypothesis stating that the use of any of our statistical illustration techniques do not provide any significant speedup in analyzing morphological properties of a collection of images and do not affect the user’s overall performance. Then, the significance value p was estimated to measure the probability of the result agreeing with the null hypothesis. For values of $p < 0.05$, the null hypothesis was rejected.

One of our first statistical measurements was estimated within the introduction of the survey. On average, we found that participants spent 3:40 minutes reading, learning, and understanding the annotation techniques. In particular, we found that on average users spent 56.18s reading about *deformation grids*, 62.62s with *likelihood volumes*, 50.56s on *line glyphs*, and 51.67s with *spherical glyphs*. After analyzing the data, we did not find any significant difference between the time used by computer scientists and non-computer scientists. Similarly, we did not find any significant difference between the time spent in each individual technique. Such findings might suggest that each illustration technique or the explanation of each approach had similar complexities.

6.1. User Preference

The first core metric captured within our user study was the users preferences. We found that on average users tend to prefer *likelihood volumes* over any other annotation technique. In particular, we found a statistically significant difference $[F(3,153) = 3.54, p < 0.01]$ between *likelihood volumes* and the other illustration techniques.

Figure 4(top) shows the overall user preferences for all the different visualization techniques. From the plot we can quickly see the significant differences between *likelihood volumes* and other annotation techniques. In addition, we can see that the preferences for *deformable grids*, *line glyphs*, and *spherical glyphs* are virtually the same. Figure 4(bottom) includes a table with the post-hoc analysis comparing the different illustration techniques. By analyzing the preferences across our two groups – engineers and non-engineers – we found similar results, but at the same time some interesting patterns.

First, a significant difference between computer and non-computer scientists was found in relation to how much they prefer *deformable grids* $[F(1,15) = 7, p < 0.01]$. On a scale from one to five, on average, computer scientists liked *deformable grids* 0.83 more than non-engineers. Figure 4(center) shows some of the results when comparing the user’s preferences across population. Another clear pattern is that both populations seem to prefer *likelihood volumes* over any of the other illustration techniques with a significance value of $p < 0.001$. Finally, from our results we can

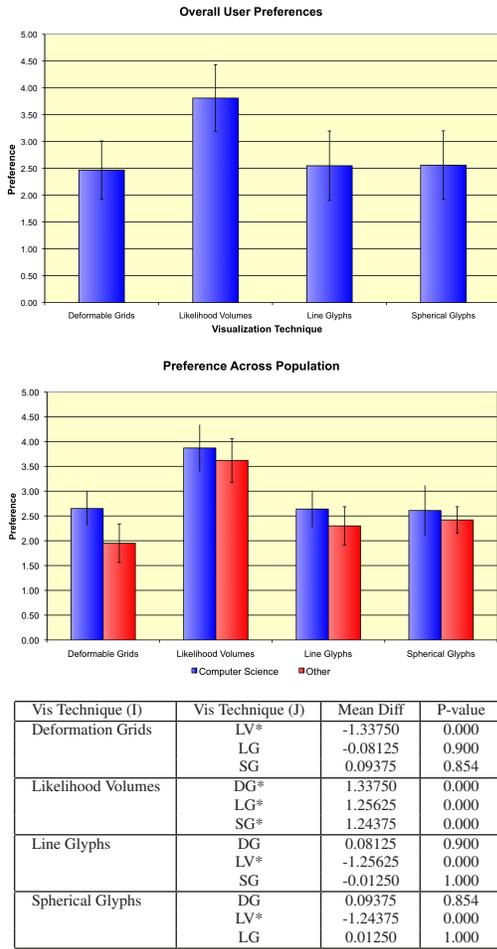


Figure 4: (top) Plot of the average user preferences and the 95% confidence intervals. We found that, on average, users tend to prefer likelihood volumes. (bottom) Plot of the user preferences across population. We found a significant difference between how much computer scientists prefer deformable grids over non-engineers subjects.

also see that computer scientists tend to like each technique more than those subjects in the other group.

We also wanted to measure how the user’s preferences change when exploring synthetic data versus medical images. Figure 5 shows some of our results. When medical images were being analyzed, there was a significant improvement [F(1,69) = 10, p<0.01] in favor of *spherical glyphs*. Among the other visualization techniques the preferences remained about the same.

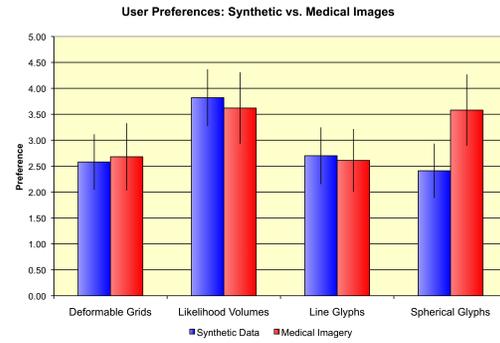


Figure 5: Plots of the user preferences when analyzing synthetic data versus medical images. We found that spherical glyphs are promising in the medical imaging domain. Specifically, from a significance test we found a P-value of 0.001 when comparing the user preferences between synthetic and medical images.

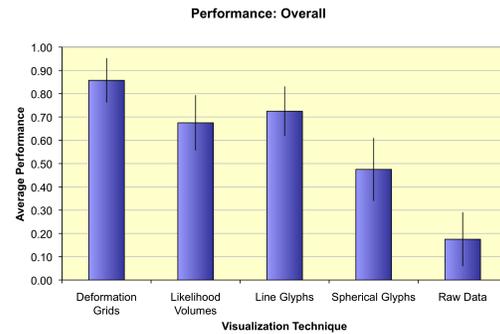


Figure 6: Plot of the overall user performance. We found that, on average, users are more accurate when exploring and analyzing statistical deformation properties with deformation grids.

6.2. User Performance

In the second section of the survey, we wanted to measure performance by capturing the user’s accuracy. In our experiments, the accuracy was defined as the average number of correct answers per user per technique.

To more effectively measure the benefits of each individual annotation technique, in this section the analysis and exploration of *raw data* was included as one technique. The inclusion of raw data helps with the overall analysis of how well statistical annotation techniques can assist during the examination of large collections of images.

Figure 6 shows the overall user performances for each individual technique. We can see that on average users performed much better with *deformation grids* than with any other annotation technique. On average we found that users were 63.3% more accurate with *deformation grids* than analyzing the raw data, 38.3% more effective than when using *spherical glyphs*, 13.3% better than using *likelihood vol-*

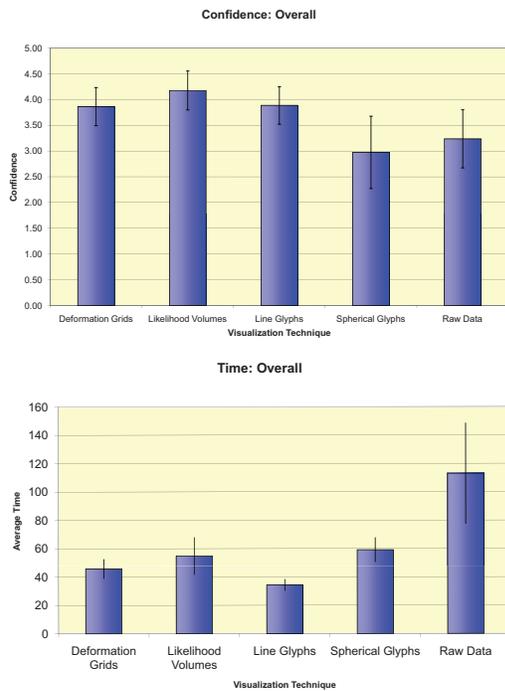


Figure 7: (left) Plot illustrating the average user confidence per annotation technique. (right) Plot illustrating the average time required to answer questions about each annotation technique. Note that analyzing raw data was clearly the technique that required the most effort to answer.

umes, and 8.3% more precise than using *line glyphs*. From the results we can see that on average there is a significant performance benefit when using *deformation grids* over *likelihood volumes*, *spherical glyphs*, and raw data. For example, a significant difference between *deformation grids* and *likelihood volumes* was found [$F(1,130) = 7, p < 0.04$]. For a complete set of results and post-hoc analysis, please see the supplemental material.

We further analyzed the overall performance across population. From the results we found that regardless the user's background, subjects performed much better with *deformation grids* than any other technique. From our experiments across population, we also found that the top three techniques to analyze morphological properties of a collection of images were similar for both groups. First, *Deformation grids*, then *Line glyphs*, and then *Likelihood volumes*.

6.3. Confidence

For each question, the users' confidence level with their answer was also captured. Figure 7(top) shows some of our results. On average, users were more confident with questions involving *likelihood volumes* and least confident with questions involving *spherical glyphs* and raw data.

When comparing the confidence level across population, we found similar patterns. First, all subjects were more confident with their answers when using *likelihood volumes*. Second, when comparing both groups, non-engineers were slightly more confident in all their answers that did not involve *deformation grids*. From our results we can see that the confidence level was somewhat biased towards their preferences. In particular, participants always felt more confident using the technique they preferred the most.

6.4. Time

The third metric captured throughout the survey was the time to answer each question. With the time results, we were able to analyze which questions and/or techniques seem to take more time to interpret and appear to have different levels of difficulty. Figure 7(bottom) shows some of our results. First, on average the amount of time required to analyze a set of images and determine their morphological deformation properties took a significant amount of time when compared with any other statistical annotation technique. Second, *line glyphs* seem to be the annotation technique where users were able to identify the statistical deformation properties the quickest, closely followed by *deformation grids*.

We found that on average users spent 45.72s on each question about *deformation grids*, 54.80s with problems regarding *likelihood volumes*, 34.33s with *line glyphs*, 59.11 on *spherical glyphs*, and 112.95 analyzing a single question with raw images.

An interesting pattern found when comparing populations was with *deformable grids*. Non-computer scientists used on average 31.4s to answer a question related to *deformable grids* while computer scientists used 50.5s. If we relate those times with the user's preferences and accuracy, we can see that non-engineers were able to more accurately and quickly pick the correct solution for an annotation technique that they did not prefer at all. This underscores the possible benefits and overall effectiveness of *deformation grids*.

7. Discussion

In general, our user study was very effective at illuminating the possible advantages and limitations of each individual annotation technique and providing a significant amount of information about the impact of each annotation technique in different subjects.

In Section 6.1 we saw the differences in user's preferences across population. In particular, in Figure 4(center) we saw a significant difference between computer and non-computer scientist in relation to how much they prefer *deformable grids*. This inclination to prefer grids might be the result of an engineering background. Engineers from many different fields frequently employ grids as a frame of reference to assist during the analysis and interpretation of 2D/3D plots [Smi06], bioengineers often use grids to show registration results [GKT*08], graphic researchers regularly use 3D

cubes and grids to better emphasize perspective, depth, and positioning [SAG*05]. We believe that such a background in engineering made computer scientists prefer, understand, and appreciate grids more than our non-engineering group. In addition, we saw that regardless the user's background, *likelihood volumes* was the preferred illustration techniques. This inclination to prefer *likelihood volumes* might be related to the visual results commonly produced by that specific annotation technique.

When comparing the user's preferences between synthetic and real-world images we found a clear tendency to prefer *spherical glyphs* when analyzing medical images (See Figure 5). This might emphasize the importance of not occluding internal structures in medical images. Our results suggest that the differences in user preferences between synthetic and medical images might be directly related to how well the annotation technique preserves internal structures. Since our synthetic data was mostly composed of 3D shapes, the internal structures were not significant when compared to medical images of the brain or other anatomical regions. In further analysis, we found that regardless the subject's background, most users preferred *spherical glyphs* in medical images as much as they liked *likelihood volumes*.

Section 6.2 showed the performance results. From our results we found that there was an statistically significant difference between using *deformation grids* over any of the other methods. If we compare these results with the users' preference, we can observe some interesting patterns. First, in Figure 4(bottom) we saw that regardless the subject's background, *deformation grids* were the least preferred annotation technique and *likelihood volumes* were the overall favorite method. However, by comparing their preference with their performance, we almost can see an inverse correlation. On average, subjects performed much better with *deformable grids* than *likelihood volumes*.

An interesting trend observed in our results was that the least favorite annotation technique becomes the method with which every subject performs the best. Second, if we just consider *likelihood volumes*, *line glyphs*, and *deformation grids*, the order happens to be the opposite when comparing preferences and performance. These findings may be related to the effect of visual appeal versus usability. In particular, over the years, researchers have noticed that techniques that are more appealing to the human eyes are not necessarily the most usable approaches. For instance, recent experiments in human-computer interaction (HCI) have found that the perception of the system's aesthetic affect the users' evaluations of usability [BBMT06]. In addition, previous work has also found that subjective evaluations of usability and aesthetics are highly correlated [TKI00]. Based on those facts, we believe that the significant inclination toward favoring *likelihood volumes* was related to the aesthetics aspects of that technique.

After the third section of the survey, we asked addi-

tional preference questions to capture preference changes over time. We found that by the end of the survey the preference order became (a) *likelihood volumes*, (b) *deformation grids*, (c) *spherical glyphs*, and (d) *line glyphs*. In other words, we found that by considering experience as a variable and comparing how the user's preferences changed after more understanding of the techniques, the user preferences for *spherical glyphs* and *deformation grids* significantly increased.

As part of the extra preference questions, we modified our glyphs annotation techniques to also include color. Now variability measurements were not only shown by the size of the glyph, but also by color. In this experiment, we found a significant increase in propensity to favor *spherical glyphs*. The average rank of *spherical glyphs* with colors became 4.1/5.0. Since this evident increase in preference towards colored spheres was not clear before our survey, the user study did not include enough questions to provide a definite answer about the effectiveness of color-mapping in statistical annotation techniques. However, based on the data gathered during the last portion of the survey, it appears color plays a crucial role on annotation of statistical deformation models.

We also compared and analyzed the correlation between the time to answer a question and the accuracy of the answer. We found that on average, there is no significant correlation and the more time users spent analyzing the questions did not translate into more accurate answers.

8. Conclusions

Our user study was very effective at providing specific information about our statistical annotation techniques. Overall, we found that procedural annotation techniques such as *deformation grids* can provide – within a single image – good insights about the statistical morphological properties of a group of images. In addition, when dealing with 2D/3D shapes, we found that illustration techniques such as *likelihood volumes* present a good method to show morphological changes. The user study also served as a tool to demonstrate once again that the most appealing visualization techniques are not necessarily the most effective methods.

There are always an infinite number of factors we can measure, test, and compare. Part of our future work includes an in-depth analysis of annotation techniques for 2D versus 3D images, a comprehensive study of the effect of colors in statistical annotation techniques, and a thorough evaluation about how glyph attributes (size, shape, position, and sampling) can affect the user's preferences and performance.

9. Acknowledgments

The authors would like to thank the anonymous reviewers for their comments and suggestions. We also would like to thank all the participants that took valuable time to complete our user study.

References

- [ALV08] ALBRECHT T., LUTHI M., VETTER T.: A statistical deformation prior for non-rigid image and shape registration. *IEEE Computer Vision and Pattern Recognition Conference (CVPR)* (2008), 1–8. 1, 3
- [BBMT06] BEN-BASSAT T., MEYER J., TRACTINSKY N.: Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Transaction of Computer and Human Interaction* 13, 2 (2006), 210–234. 9
- [BH07] BAIR A., HOUSE D.: A grid with a view: Optimal texturing for perception of layered surface shape. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1656–1663. 2
- [CR00] CEDILNIK A., RHEINGANS P.: Procedural annotation of uncertain information. In *VIS '00: Proceedings of the conference on Visualization '00* (2000), pp. 77–83. 2, 4
- [CWP*01] CHUNG M. K., WORSLEY K. J., PAUS T., CHERIF C., COLLINS D. L., GIEDD J. N., RAPOPORT J. L., EVANS A. C.: A unified statistical approach to deformation-based morphometry. *NeuroImage* 14, 3 (2001), 595 – 606. 1
- [GKT*08] GLOCKER B., KOMODAKIS N., TZIRITAS G., NAVAB N., PARAGIOS N.: Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis* 12, 6 (2008), 731 – 741. 8
- [GR04] GRIGORYAN G., RHEINGANS P.: Point-based probabilistic surfaces to show surface uncertainty. In *TVCG* (2004), vol. 5, pp. 564–573. 2
- [HIL*05] HOUSE D., INTERRANTE V., LAIDLAW D., TAYLOR R., WARE C.: Design and evaluation in visualization research. *Visualization Conference, IEEE 0* (2005), 117. 2
- [HMBG01] HAUSER H., MROZ L., BISCHI G. I., GRÖLLER M. E.: Two-level volume rendering. *IEEE Transactions on Visualization and Computer Graphics* 7, 3 (2001), 242–252. 4
- [KHI*03] KOSARA R., HEALEY C., INTERRANTE V., LAIDLAW D., WARE C.: User studies: Why, how, and when? In *IEEE Computer Graphics and Applications* (2003), vol. 23, pp. 20–25. 2
- [Kin04] KINDLMANN G.: Superquadric tensor glyphs. In *Proceeding of The Joint Eurographics - IEEE TCVG Symposium on Visualization* (May 2004), pp. 147–154. 2
- [KUS*05] KNISS J. M., UITERT R., STEPHENS A., LI G., TADIZEN T.: Statistically quantitative volume visualization. In *IEEE Visualization Conference 2005* (2005), pp. 287–294. 2
- [Lev60] LEVENE H.: Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (1960), pp. 278–292. 6
- [LKJ*05] LAIDLAW D. H., KIRBY M., JACKSON C., DAVIDSON J. S., MILLER T., DASILVA M., WARREN W., TARR M.: Comparing 2d vector field visualization methods: A user study. *Transactions on Visualization and Computer Graphics* 11, 1 (2005), 59–70. 2
- [LKJ*06] LEOW A. D., KLUNDER A. D., JR. C. R. J., TOGA A. W., DALE A. M., BERNSTEIN M. A., BRITSON P. J., GUNTER J. L., WARD C. P., WHITWELL J. L., BOROWSKI B. J., FLEISHER A. S., FOX N. C., HARVEY D., KORNAK J., SCHUFF N., STUDHOLME C., ALEXANDER G. E., WEINER M. W., THOMPSON P. M.: Longitudinal stability of mri for mapping brain change using tensor-based morphometry. *NeuroImage* 31, 2 (2006), 627 – 640. 1
- [LP07] LJUNG P., PERSSON A.: Uncertainty visualization in medical volume rendering using probabilistic animation. In *IEEE Transactions on Visualization and Computer Graphics* (2007), vol. 13, pp. 1648–1655. 2
- [Nor06] NORTH C.: Toward measuring visualization insight. *IEEE Comput. Graph. Appl.* 26, 3 (2006), 6–9. 2
- [NS00] NORTH C., SHNEIDERMAN B.: Snap-together visualization: Can users construct and operate coordinated views? *International Journal of Human-Computer Studies* 53, 5 (2000), 715–739. 2
- [PCS05] P. COLEMAN SAUNDERS VICTORIA INTERRANTE S. C. G.: Pointillist and glyph-based visualization of nanoparticles in formation. In *IEEE EuroVis Conference* (2005), pp. 169–176. 2
- [PKR06] POTTER K., KNISS J., RIESENFELD R.: Visual summary statistics. In *Tech report UUCS-07-004, University of Utah* (2006). 2
- [PWL97] PANG A. T., WITTENBRINK C. M., LODHA S. K.: Approaches to uncertainty visualization. *The Visual Computer* 13 (1997), 370–390. 2, 4
- [RFS03] RUECKERT D., FRANGI A., SCHNABEL J.: Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Transactions of Medical Imaging* 22 (2003), 1014 – 1025. 1, 3
- [SAG*05] SHIRLEY P., ASHIKHMIN M., GLEICHER M., MARSCHNER S., REINHARD E., SUNG K., THOMPSON W., WILLEMSSEN P.: Fundamentals of computer graphics. In *ISBN-10: 1568812698* (2005), no. 2. 9
- [Smi06] SMITH S. T.: *MATLAB Advanced GUI Development*, 1 ed. Dog Ear Publishing, LLC, 2006. 8
- [TKI00] TRACTINSKY N., KATZ A., IKAR D.: What is beautiful is usable. *Interacting with Computers* 13 (2000), 127–145. 9
- [TM04] TORY M., MÖLLER T.: Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics* 10, 1 (2004), 72–84. 2
- [WPL96] WITTENBRINK C., PANG A., LODHA S.: Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics* 2, 3 (1996), 266–279. 2