

CareerVis: Hierarchical Visualization of Career Pathway Data

Mingran Li¹
Wenjie Wu¹
Junhan Zhao¹
Keyuan Zhou¹
David Perkis²
Timothy N. Bond²
Kevin Mumford²
David Hummels²
Yingjie Victor Chen^{1,2}

¹Department of Computer Graphics Technology, Purdue University

²Krannert School of Management, Purdue University

Editor:
Mike Potel
potel@wildcrest.com

We present our CareerVis system, an interactive visualization tool to aid career education for high school and freshman college students. In addition to its practical use, we believe our design approach has potential to inspire the design community to develop simple visualizations that convey complex information to novice users.

To help students prepare for success in college, career, and life, education stakeholders have a long-term commitment to developing curriculum that allows students to explore their college and career options as well as their aptitudes and employability (www.doe.in.gov/sites/default/files/standards/cte-family-and-consumer-sciences/cf-busfacs-preparingcc_7-11-14.pdf). Studies have shown that the right kinds of education can help people ease their transition into the job market.¹ Stu-

dents should be able to plan for college and career pathways that are suitable for their interests, abilities, and lifelong goals.²

Within this context, the community needs efficient tools to help students better prepare for their careers after graduation.³ Based on a synthesized dataset from Purdue graduates' job placement survey data and a national survey database (www.onetonline.org; www.mynextmove.org), we developed CareerVis, a visualization system aimed to help young students comprehend the broad range of educational and occupational paths as part of the college-career selection process.

In order to improve the efficiency of decisions made by students, parents, and other career education stakeholders, we need to anticipate the following frequently-asked questions:

- 1) Which major should I choose? Which occupation should I pursue?
- 2) What majors can help me pursue this occupation?
- 3) What occupations am I qualified for with this major?
- 4) What are the characteristics of these majors and occupations?

THE DATA AND MESSAGE

Our underlying dataset contained flow information with hierarchical structure and multidimensional characteristics, which could be decomposed into several typical data structures in information visualization design.⁴ Particularly, the dataset was composed of college majors, occupations, flows of students from majors to their first jobs, 12 numeric measurements for majors (e.g., GPA and SAT scores), and 12 measurements for occupations (e.g., salaries and future trends in globalization and automation). The dataset presented the following challenges on visualization:

- 1) three different types of data formed a more complex data structure;
- 2) the relatively large amount of data;
- 3) the data would be presented for the general public, who have little experience with reading and understanding visualization applications; and
- 4) the visualization would also be published on a relatively small screen (e.g., a tablet or smartphone).

The data is inherently hierarchical. Purdue University's West Lafayette campus houses 145 departments within the 10 colleges: Agriculture, Education, Engineering, Health and Human Sciences, Liberal Arts, Management, Pharmacy, Science, Technology, and Veterinary Medicine. Students' occupations were aggregated into 130 specific job positions and put into occupation groups by the standard occupational classification system (www.bls.gov/soc). Moreover, the data contains proportions of the student body enrolled in majors or landed in a different job.

There are multiple pathways for students to pursue their ideal occupations from majors. For instance, among students who received accounting degrees, 70% secured accountant and auditor positions, while 17% worked as financial analysts, and 1%–3% worked in 8 other occupations. Vice versa, other than two majors (Accounting and General Management) under Management, about 10% of accounts come from majors under six different colleges which range from Agriculture to Science. There are about one thousand possible major-to-occupation pathways.

College majors and occupations have many numerical measurements that can offer students insight into the requirements of certain majors (e.g., GPA, SAT scores required for enrollment), important measurements for jobs (e.g., salary and work hours), characteristics of certain majors and occupations (e.g., percentage of Indiana students, percentage of domestic/international students, and diversity status of minorities), and future trends that could be affected by automation and globalization. These characteristics are complementary descriptions of majors and jobs necessary to guide students' decision making. People may find certain jobs are better suited for family-oriented employees since their percentage of married workers is higher than other professions (e.g., engineering and construction). Some job opportunities may increase (or decrease) with the development of automation and globalization. Some characteristics have single percentage values (e.g., the percentage of the workforce from various ethnic groups), while others have percentile values of 10%, 25%, median, 75%, and 90%, such as salary, GPA, and SAT scores.

DESIGN EXPLORATION AND ITERATIONS

Our team conducted several iterative designs and involved users in the design processes. From more than ten design ideas [e.g., Figure 1(a)], we selected the one with the most simple and intuitive form to develop. Figure 1(b) shows our first formal design in horizontal layout. Characters are visualized as heat maps. Flows are presented in a hierarchical Sankey diagram. Our testing showed that the majority of users fully comprehended the hierarchies between the colleges and majors, as well as the

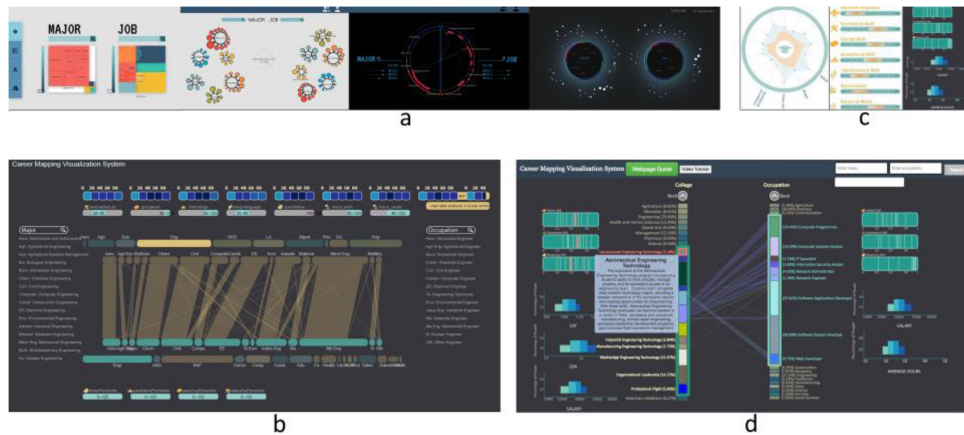


Figure 1. CareerVis's early design concepts and iteration: (a) several concept designs; (b) horizontal layout; (c) better solutions for characteristic plots of vertical design; (d) vertical layout.

occupational groups and occupations. However, the horizontal design presented several problems such as the interface layout generated unnecessary information overlap.

To better utilize screen space and represent characteristics, we developed another version. The main design idea of hierarchical flow remains but was rotated into a vertical presentation. To improve the design of side characteristics, we brainstormed many solutions [Figure 1(c)]. Ultimately, we selected a relatively intuitive option that featured the direct relationship between the line and value to show the characteristics of verbal, quantitative, and reasoning skills [Figure 1(d)]. The straight line would be highlighted with bright red when hovering on a particular major or occupation. Additionally, we presented other characteristics by bar graphs. Based on interviews with 64 participants, 90% of users were more satisfied with the vertical design. We further enhanced this design to address some remaining problems with the implementation of interactions, mainly the lack of a comparison function and the fact that once a user clicked on a college or an occupation group, the block expanded suddenly, which caused the user to lose visual momentum due to a sudden change of the visualization.

VISUAL COMPONENTS

Our resulting CareerVis user interface provided multiple views for data exploration [Figure (2)]. The central flow view displayed the hierarchical levels of colleges, majors, occupational groups, and occupations, as well as the one-to-one relationship between major and occupation. Each particular major or occupation featured a complete description. All relevant characteristics of majors and occupations were presented by scatter plots, box plots, and bar charts.⁵ In the top view, the system provided a guide and a search function. The system is developed using d3 (d3js.org).

Figure 2(a) shows the breadth of occupations chosen by students after graduating from college. In the central section, our team combined three essential visual components:

- 1) the rectangular blocks of the Purdue colleges and majors on the left side, where lengths of blocks represent the number of students that have graduated;
- 2) the similar visual element of the occupational groups and occupations on the right side, where lengths represented the number of students in the occupational group; and
- 3) the connection paths in the central region.



Figure 2. Overview of CareerVis system with the following six sections: (a) career flow; (b) major quantitative/verbal; (c) occupation quantitative/verbal; (d) major characteristics; (e) occupation characteristics; (f) accordion of characteristic plots.

The paths that connected both sides show the percentages of students who graduated with a particular college major(s) and obtained a specific job position(s). The college/major and occupation/occupational group were embodied in the two content-rich side bars. Clicking on a college expand the college into majors.

Characteristics plots [Figures 2(b), 2(c), 2(d), and 2(e)] provide users with detailed information about the colleges, majors, occupational groups, and occupations. For the verbal and quantitative scores, we use scatter plots, one for the college side and another for the occupational side. Dots in the plot represent majors or occupations. The color of a dot is based on the major/occupation's requirement of verbal and quantitative skills. For characteristics like salary, SAT score, and GPA, we took advantage of their 10th, 25th, 50th, 75th, and 90th percentile data and use the box plot for their visualization. Users can not only compare the differences between majors and occupations, but they can also view the data distribution within the major or occupation.

The accordion⁶ [Figure 2(f)] was introduced to represent many characteristics of majors and occupations. With the accordion, the system can contain numerous characteristics' charts within a relatively small space. Users can expand the characteristics they want and fold those in which they have no interest. An ordinary PC or laptop screen can hold four characteristics charts at the same time. The most important characteristics are opened by default on top of the list when users enter the system. If a user opens several charts that extended beyond the screen, they can scroll down the screen. The system automatically adjusts the position of the central flow visualization part and keep it centered in the current window.

Color is used to encode the two basic skills—quantitative skills and verbal skills—required by each occupation and major. If one major/occupation requires more quantitative skills, its color is

bluer. Yellow means the major/occupation requires balanced quantitative and verbal skills, whereas red means the major/occupation requires more verbal skills. The saturation of color represents the strength of skills these majors/occupations required. The quantitative/verbal Cartesian coordinate system is mapped into a hue (blue to yellow to red) saturation polar coordinate system. The radius (saturation) is computed by the distance of the quantitative/verbal (X/Y) to the origin (0, 0). The hue is computed by the angle of the quantitative/verbal point to the X-axis (verbal). Colors are featured consistently across all sections of the quantitative/verbal scatter plot, characteristic graphs, and center major/occupation flows.

INTERACTION AND ANIMATION

Brushing and Linking

Brushing and linking allow users to locate a major or occupation from the side charts and verify the corresponding values.^{7,8} The system's brushing and linking occur when users hover over a bar in the center diagram, and the corresponding side chart elements are highlighted. When users hover over the College of Engineering in the center diagram, for example, the circle of majors in the corresponding scatter plot are highlighted by a thin black stroke and other circles fade [Figure 3(a)], and the bars of the College of Engineering are highlighted in every opened characteristic chart by a gray line. The values of each characteristic are shown at the top of each gray line [Figure 3(b)], which allows users to easily confirm the exact values. Brushing and linking are active within both levels of the chart, which means that when hovering over a college, occupation, or major with an open center chart, the corresponding elements are all be highlighted in the charts of the corresponding sides.



Figure 3. Brushing and linking of (a) scatter plot; (b) other characteristics charts.

Focus

A focus stage was added to the bars for when users hover over the parallel sets. When a bar is focused, the transparency of the paths linked to this bar increases, and the transparency of the other paths decreases to the extent that users can easily distinguish the focused paths from other paths and distributions. There are some differences in the interactions between the first and second hierarchies, depending on which one becomes focused, but the percentage of each path will be displayed.

Contexts and Details

The system constantly locates users and compares the current focus with related information.⁸ There are two categories of context maintenance. The first is used in the flow path of the Sankey diagram [Figure 4(a)] and the scatter plot [Figure 4(b)]; the second is applied to the bar set of the Sankey diagram [Figure 4(c)], box plots, and bar charts of the second-level display [Figure 4(d)]. For the flow path, as mentioned above, when a bar set is focused, the transparency of the connected paths increases, and the other paths become transparent. We maintain the paths as light background context elements. We use the same strategy of highlighting the selected circles and fade the others within the scatterplot. As for the two levels, we keep the first level elements displayed when the second level is opened. In the center Sankey diagram, the unselected first-level bars are shrunk and faded. In the box plots and bar charts, the unselected first-level bars are shrunk and moved to the left or right based on their position relative to the selected bar.



Figure 4. Zoom in for contextual details: (a) central flow of majors in one college to occupations in a group; (b) scatter plot highlighting focused majors and occupations; (c) context of other college and occupation groups; (d) detailed characteristics of occupations within the context of all other groups.

Animated Transition

The system's animated transitions help the users understand the relationships between the two levels.⁹ The rich interaction causes the system to change its visual elements significantly. To maintain the cognitive coupling of the user with the system, we reinforce visual momentum by incorporating animated transitions in both the center diagram and the characteristic charts.¹⁰ For the center Sankey diagram, when a college bar is opened, the college bar gradually expands to a settled length, and the second college bar levels, which are the majors within this college, gradually expand and occupy the college bar area. At the same time, other college bars shrink to the same smaller size for readability [Figure 5(a)]. The occupation side applies the same rule. The transitional animations of the characteristic charts follow the same rule: the selected bar gradually expands and disappears, the corresponding second level bars gradually appear and expand, and the other first level bars shrink [Figure 5(b)]. All of these animations take 0.5 s, a value selected after user testing of different animation lengths.

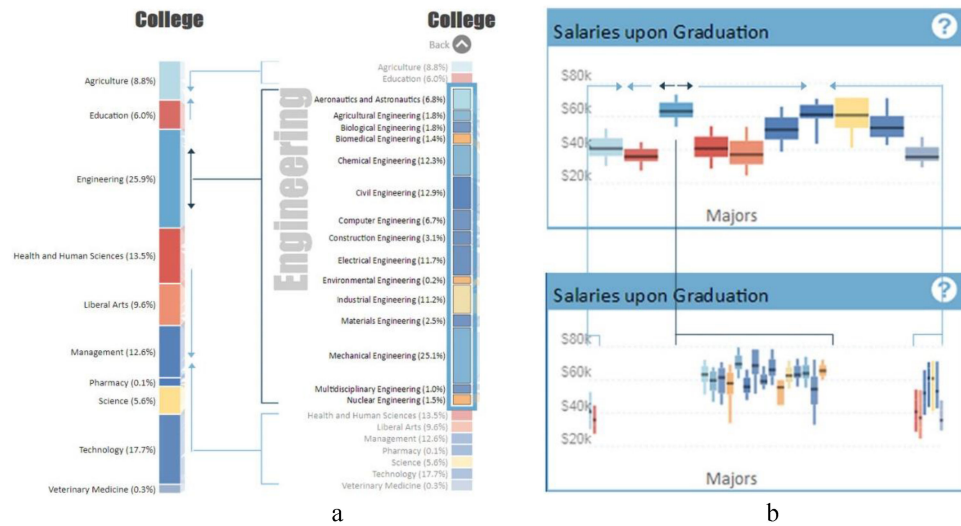


Figure 5. Animation design context details of the (a) central flow and (b) two-side characteristics.

General Information Query With the Central Flow

Assume a user is a student from mechanical engineering (ME) and they want to find out what occupations they may have in the future. The user should first open the college of Engineering, mouse-over to ME to see connections to occupations several occupation groups [Figure 6(a)], then open the major occupation group of Engineers [Figure 6(b)] to see the most

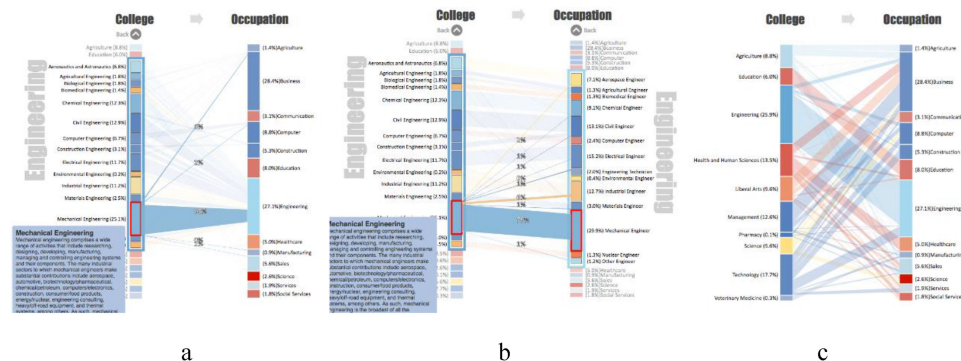


Figure 6. Flow of interaction in the center graph.

frequent occupations from ME. The user can then close both sides to return to the initial flow view [Figure 6(c)].

Characteristics

Assuming a user is interested in the engineering major, they can compare the differences in verbal and quantitative skills required for the major. When they open the College of Engineering chart, several converging points, 0.65–0.9 verbal and 0.6–0.9 quantitative, are highlighted. When hovering on ME, the dots of 0.9 verbal and 0.9 quantitative are highlighted with a red stroke [Figure 7(a)]. Compared with ME, computer engineering requires on average 0.65 verbal and 0.8 quantitative [Figure 7(b)].

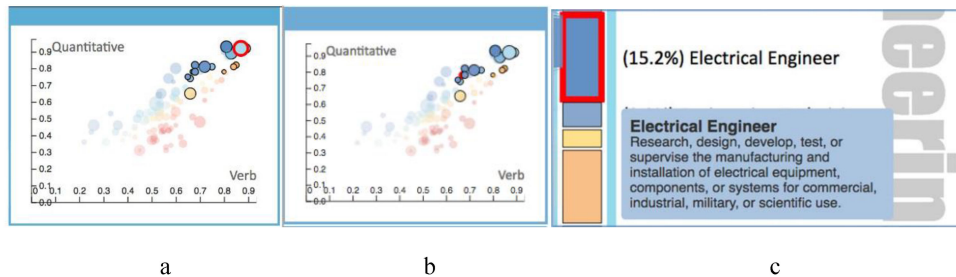


Figure 7. Hypothetical user's CareerVis interface flow with the corresponding scatter plots. Explore and compare the values of quantitative and verbal skills between different college majors/ occupations (a, b, and c).

Moreover, when the user moves on to electrical engineering, they find that this major requires 0.75 verbal and 0.8 quantitative. They decide the best possible occupation for their major is Electrical Engineer based on these statistics [Figure 7(c)]. Furthermore, the user proceeds to the side box salary plot and notices that the median earnings for that occupation are \$65k [Figure 8(a)]. They learn the average hours of work per week by an electrical engineer is 43 [Figure 8(b)], and the occupation's value of automation sensitivity is 10 [Figure 8(c)], which means it is very sensitive to the development of automation. They can open up different characteristic tabs to develop their overall knowledge of the pathway. Instead of clicking on a college/occupation group in the center flow bar to see all detailed majors and occupations, the user can click on a box or a bar in the characteristic graphs to open up the college/occupation in the center graph.

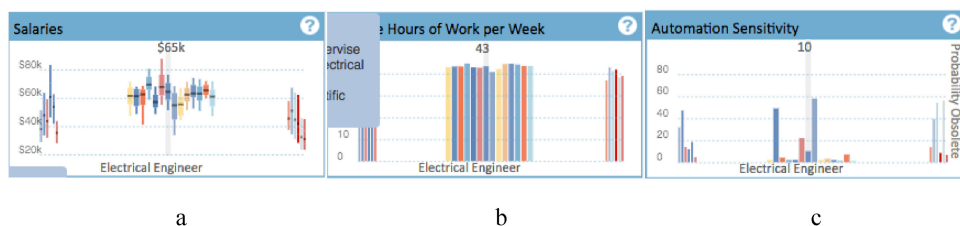


Figure 8. Hypothetical user's CareerVis interface flow, including the corresponding bar graphs and box plots. Characteristics of the focused element marked by gray bars (a, b, and c).

USER FEEDBACK

In our usability testing, we designed quantitative and qualitative questions to compare participants' understandings and expectations toward the relationship between college majors and occupations. The number of the valid participants was 68, all of whom were first-year students who came from Purdue University. We recorded their mouse activities during the experiments and only nominated the participants who spent more than one minute actively in our system, while the average spending time was more than 30 min. As the result, the overall compatibility

of participants' selected majors and occupations increased from 79% to 91% after using our system. Additionally, 13 changed their primary major selection, and 30 changed their primary occupation selection. They further provided feedback about how they evaluated this application by ranking from best to worst various aspects of this application: attractiveness, clearness, colorfulness, helpfulness, difficulties, effectiveness, and efficiency. Their responses indicated that this application received the highest scores in being helpful, colorful, and attractive, with difficulty and efficiency being areas for improvement.

In 2017, the Indiana Department of Education proposed a set of learning objectives for the curriculum to prepare students for college and careers (www.doe.in.gov/sites/default/files/standards/cf-bus-facs-pcc-01-2016.pdf). Purdue curriculum specialists examined our CareerVis tool and found that the tool could help achieve these learning objectives and that the tool would be useful to integrate into the course curricula to improve college and career teaching and learning.

CONCLUSION

Based on recent years' job placement data from Purdue students, we designed and developed our visualization system, CareerVis, to represent the vast choices of education and pathways to the occupation. Using innovative yet simple visual forms, we believe we solved two big challenges in the visualization. First, our system is able to deal effectively with hundreds of majors and occupations, and a thousand possible pathways. Second, our system has proven to be easily comprehensible by a general audience, such as high school students, parents, and educators. Collaborating with experts in education, we are now working toward integrating this system with curriculum to address the Indiana Department of Education's high school academic standards for Preparing for College and Careers.

ACKNOWLEDGMENTS

This work was supported by a grant to Purdue University from the Lilly Endowment Inc. The grant is part of the Lilly Endowment's "Round III: Initiative to Promote Opportunities through Educational Collaborations." Access the CareerVis visualization system at: va.tech.purdue.edu/careerVis.

REFERENCES

1. F. A. Levy, *The New Division of Labor: How Computers are Creating the Next Job Market*. Princeton, NJ, USA: Princeton Univ. Press, 2005.
2. N. J. Evans, D. S. Forney, F. M. Guido, L. D. Patton, and K. A. Renn, *Student Development in College: Theory, Research, and Practice*. Hoboken, NJ, USA: Wiley, 2009.
3. W. A. Anderson, "Important events in career counseling: Client and counselor descriptions," *Career Dev. Quarter.*, vol. 48, pp. 251–263, 2000.
4. B. Lee, C. Plaisant, C. S. Parr, J. D. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proc. AVI Workshop Beyond Time Errors: Novel Eval. Methods Inf. Vis.*, May 2006, pp. 1–5.
5. J. Heer, M. Bostock, and V. Ogievetsky, "A tour through the visualization zoo," *Queue*, vol. 8, no. 5, p. 20, 2010.
6. A. A. Cooper, *About Face: The Essentials of Interaction Design*. Hoboken, NJ, USA: Wiley, 2014.
7. A. Buja, D. Cook, and D. F. Swayne, "Interactive high-dimensional data visualization," *J. Comput. Graph. Statist.*, vol. 5, no. 1, pp. 78–99, 1996.
8. J. S. Yi, Y. ah Kang, and J. Stasko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1224–1231, Nov./Dec. 2007.
9. J. Stasko and E. Zhang, "Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations," in *Proc. IEEE Symp. Inf. Vis.*, 2000, pp. 57–65.
10. D. D. Woods, "Visual momentum: A concept to improve the cognitive coupling of person and computer," *Int. J. Man-Mach. Studies*, vol. 21, no. 3, pp. 229–244, 1984.

ABOUT THE AUTHORS

Mingran Li is currently working toward the Ph.D. degree with the research focus on information visualization design at Purdue University, West Lafayette, IN, USA. Contact her at li1940@purdue.edu.

Wenjie Wu is a User Experience Designer. She received the Master of Science degree from Purdue University, West Lafayette, IN, USA. Contact her at wenjie.jessie.wu@gmail.com.

Junhan Zhao is currently working toward the Ph.D. degree at Polytechnic Institute, Purdue University, West Lafayette, IN, USA. Contact him at zhao835@purdue.edu.

Keyuan Zhou is a Research Assistant with the Polytechnic Institute, Purdue University, West Lafayette, IN, USA. Contact him at zhou851@purdue.edu.

David Perkis is Director of the Purdue Center for Economics, Purdue University, West Lafayette, IN, USA. Contact him at perkis@purdue.edu.

Timothy N. Bond is an Assistant Professor with the Economics Krannert School of Management, Purdue University, West Lafayette, IN, USA. Contact him at tnbond@purdue.edu.

Kevin Mumford is Director of the Purdue University Research Center in Economics (PURCE) , West Lafayette, IN, USA. Contact him at mumford@purdue.edu.

David Hummels is Dean and Professor of economics with Purdue University, West Lafayette, IN, USA. Contact him at hummelsd@purdue.edu.

Yingjie Victor Chen is an Associate Professor with the Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, USA. Contact him at victorchen@purdue.edu.

Contact department editor Mike Potel at potel@wildcrest.com.