2019

Micro-Level Analysis and Visualization of Tennis Shot Patterns with Fractal Tables

Shiraj Pokharel

Ying Zhu

Satish Puri

2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), (2019) : 97-102. DOI.

Micro-Level Analysis and Visualization of Tennis Shot Patterns with Fractal Tables

Shiraj Pokharel
Department of Computer Science, Georgia State University
Ying Zhu
Creative Media Industries Institute, Georgia State University
Satish Puri
Department of Mathematics, Statistics and Computer Science, Marquette University

Abstract:
Sports data analysis and visualization can be a useful tool for gaining insights into the games. In this paper, we present a new technique to analyze and visualize shot patterns in tennis matches. Tennis is a complicated game that involves a rich set of tactics and strategies. The current tennis data analyses are usually conducted at a high level that often fail to show useful patterns and nuances embedded in low level data. Based on a very detailed database of professional tennis matches, we have developed a system to analyze the serve and shot patterns so that a user can explore questions such as "What are the favorite patterns of this player? What are the most

effective patterns for this player?" This can help tennis experts, players, and fans gain deeper insight into the sport.

SECTION I. INTRODUCTION

Sports data analysis and visualization can help players, coaches and analysts gain insight into the game and therefore help them improve their performance. Such analysis can also help sports fans appreciate and enjoy the games on a deeper level. Therefore, most sports event broadcasters and web sites often present some statistics at the end of a match. For example, after each tennis match, a TV broadcaster or a news web site will list or visualize the total number of aces, unforced errors, winners, the serve percentages, etc. Such statistics are often displayed without explanation. High-level statistics is useful but they often fail to reveal the complexity of tennis matches. As more detailed datasets become available, we are able to conduct micro-level data analysis that reveal deeper insight into the dynamics of the match. In this paper, we present a new data analysis and visualization technique for analyzing tennis shot patterns. Shots are the basic building blocks of tennis matches. A tennis tactic is a combination of particular shots. In other words, each tennis tactic is a particular shot pattern. By analyzing and visualizing shot patterns, our system can show how different players use different tactics to try to gain advantages. Currently shot patterns are often visualized as heat maps or shot trajectory maps. Heat maps can show where the shots land on court, but they cannot differentiate various combinations of shots. They also do not differentiate long rallies and short rallies. Shot trajectory maps can potentially show shot combinations, but when many shots are displayed, such maps become crowded and hard to read. We address these problems by introducing a new visualization technique – fractal tables. A fractal table is a table that can be subdivided recursively as needed. In our method, each fractal table contains tennis points of a certain rally length. Each cell in a fractal table represents a unique combination of shots – a shot pattern. A tennis match can be visualized by placing each point in a cell of a fractal table. Because the fractal tables can be recursively subdivided, they can accommodate tennis points of any rally length and any combination. The fractal tables provide a clear and efficient technique for analyzing tennis shot patterns on a micro-level. For example, long shots and short shots are separated. If a player favors certain shot patterns, they will be clearly visible in the fractal tables because more points will be placed in particular cells. The fractal tables can show which player has advantage (or disadvantage) in certain shot patterns. Overall, this new data analysis and visualization technique can help us better understand a player’s tactical and strategical thinking during the match. We demonstrate our new technique by presenting a case analysis of a professional tennis match.

SECTION II. RELATED WORKS

There has been a good number of research literature on sports data visualization and analytics in general and tennis visual analytics in particular. For example, He and Zhu [1] proposed a data visualization that shows the progression and tactical statistics of a tennis match. Pokharel and Zhu [2] proposed models for visual analytics of performance anxiety in tennis. Other researchers [3] [4] [5] [6] [7] [8] have done some important work on statistical modeling of tennis matches. A tennis match visualization system that shows the score, point outcomes, point lengths, service information, and match videos for tennis enthusiasts and coaching staff to gain insights into match performance has been proposed in [9]. Liqun and Banks [10] proposed a technique to visualize the overall structure of the match as well as the fine details using a 2D display of translucent layers derived from Tree-Maps. Burch and Weiskopf [11] introduced techniques to visually encode the dynamics of a tennis match by using a hierarchical and layered icicle representations. They place the time axis vertically as multiple aligned scales to indicate the duration of games and points. They also used color coding to visualize additional attributes. An approach to quantify the similarity between players, which could be used for prediction and recommendation has been proposed in [12]. Probabilistic graphical models to study player behavior, which could be used to find the factors such as location and speed of the incoming shot has been developed in [13].

Previous work in tennis data analysis and visualization mainly focus on the point level data or higher. They rarely dealt with micro-level analysis such as the tactical shot patterns, which is the focus of our work. Data analysis and visualization work has been done for other sports. For example, a visualization of table-tennis matches that includes time-oriented, statistical, and tactical analyses has been proposed in [14]. Working with domain experts to present a visual analytics system for soccer data, allowing users to track the spatio-temporal changes in formation and understand how and why such changes occur is proposed in [15]. A system to analyze high-frequency position-based soccer data at various levels of detail for analysis of movement features and game events is presented in [19].

SECTION III. Data

The tennis match data is made available by Tennis Abstract [16], an open source project that provides shot-by-shot statistics of more than 5000 professional tennis matches, including the type of shot, direction of shot, depth of returns, types of errors, etc.

SECTION IV. Method

A. Fractal Table

A fractal table is a table that can be subdivided recursively as needed. Each cell in the fractal table represent a particular shot combination (e.g. serve wide -¿ backhand return -¿ forehand -¿ backhand). The longer the rally, the more subdivision is needed to accommodate larger number of possible shot combinations. Horizontally, the factual table is arranged by shot length, starting on the left with one shot (service aces), then two-shot combination (serve and return), three-shot combination, four-shot combination, etc. In this paper, we only show fractal table up to four shots because of space limits and also because about 70% of the points in professional tennis end in four or fewer shots. Therefore, a one-to-four shot fractal table can capture the majority of the points. Our fractal table can be easily extended to longer rallies if needed. In the fractal table, the one-shot space (service aces) is vertically divided by player A's serve directions: Wide(W), Body(B), and down-the-T(T). The two-shot space adds one division to the one-shot space – each service direction box is divided horizontally by player B's return types: forehand return and backhand return. The three-shot space adds one division to the two-shot space – each service return box (e.g. forehand return) is divided vertically by player A's shot type: forehand and backhand. The four-shot space adds one more division to the three-shot space – each shot box is divided vertically by player B's shot type: forehand and backhand. And so on. Therefore, player A's shot types are all divided vertically and player B's shot types are all divided horizontally. The benefit is that we can easily group player A's forehand or backhand shots because they are all in parallel rows. Similarly, we can group player B's forehand or backhand shots because they are all in parallel columns. This can help us identify shot patterns, strength, and weaknesses. Each point in a tennis match is represented as a dot in one of the boxes in the fractal table, based on the shot combination (e.g. Player A serve wide -¿ Player B backhand return -¿ Player A forehand -¿ Player B backhand) and the box division rules described above.

Fig. 1. A four shot space

B. Data Visualization

Figure 1 depicts a fractal table that can visualize points of up to four shots. The shot combination for each box is marked in white letters. In Figure 1, the three dark-orange colored boxes are for one-shot points: aces and service winners. The khaki colored boxes are for two-shot points. For example, the box labelled BB will contain points with serve to the body (B) followed by a backhand return (B). The box labelled TF will contain points with serve down the T followed by a forehand return (F). The green boxes are for three-shot points. For example, a green colored box labelled BFB will contain points with a serve to the body (B), a forehand return (F), and the server's backhand shot (B). The blue colored boxes are for four-shot points. The visualization was implemented in Javascript library D3 [18].

C. Visual Analytics Techniques

We can use the fractal table to conduct different visual analytics. For example, by analyzing the serve directions, we can see different player's serve patterns. For example, who likes to serve wide whether such serve is effective. We can also see what kind of shot patterns a particular player likes to use after serve wide, down the T, or to the body. We can see whether a player prefers a particular shot pattern at critical moments (e.g. game point, break point, etc.) Is a player more effective in playing long points (e.g. a defensive player) or playing short points (e.g. a offensive player)? We can highlight certain shot patterns where one player needs to run a lot (e.g. a shot pattern like WFFBFFFBFF). Therefore, we can see if a player is good at hitting on the run. We can also highlight certain shot patterns where one player dictates the point with forehand (e.g. a shot pattern like WFFBFFFBFF) and see who is the dominate player in the game.

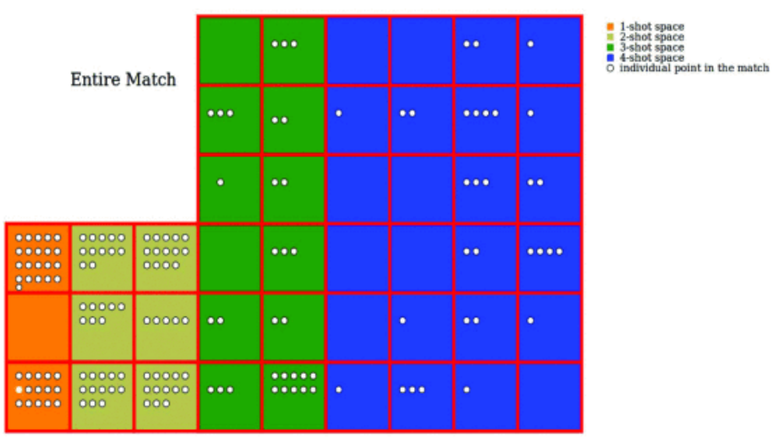

Fig. 2. Points Plotted for the entire match.

SECTION V. Case Studies

In this section we demonstrate our method using the match data from the 2017 US Open semi-final between ATP players Kevin Anderson and Pablo Carreno Busta (henceforth referred to as KA and PCB respectively), where KA defeated PCB 4–6, 7–5, 6–3, 6–4.

A. Full Match Analysis

The two players played 247 points, including 36 one-shot points, 65 two-shot points, 31 three-shot points, and 31 four-shot points. Therefore, 163 shots, or 66% of the total shots, are within the one to four-shot range. In our analysis, we focus on the first four shots in tennis for several reasons. The majority of the shots are four shots or fewer. Some leading tennis analysts, such as Craig O'Shannessy (tennis analysts for the New York Times and world 1 Novak Djokovic), believe that the first four shots are the most important in tennis. We also want to keep our fractal tables relatively small so they can fit the space of this paper. Figure 2 shows all the one to four-shot points grouped by shot patterns. Each white dot represents a point. The data visualization reveals some patterns in the player's shot selections. In the one-shot space (i.e. aces or serve faults), we see that more points were served down the T than wide, and there was no serve to the body. This indicates that serving down the T is more likely to produce an ace. In comparison, the two-shot space contains almost equal number of forehand and backhand return of serves. The two-shot space has the most points than any other shot pattern, perhaps indicating the effectiveness of the players' serves and returns. In the three and four-shot spaces, the majority of the points were played with a backhand return of serve. For example, in the four-shot space, 74% of the points were played with a backhand return of serve. This indicates that serving to the backhand tends to result in short rallies.

B. Set By Set Analysis

Figures 3 to 6 show the shot patterns broken down by sets. From these visualizations, we can observe some interesting patterns. First, no ace was generated from serving to the body. Also in general, fewer points were played from serving to the body. This indicates that both players prefer to serving wide or to-the-T. In the one-shot space, we see that in the fourth set, there are significantly fewer one-shot points than previous sets, perhaps indicating the players were tired and their serves were less powerful than earlier in the match. There are four serves to the T and only one serve wide, suggesting that serving to the T was still effective at this stage of the match, perhaps because the net is lower in the middle. In the second set, we can see a remarkable pattern in the three-shot space. There are five points played in the WBF (Serve Wide from ad side -¿ Backhand return -¿ Forehand) combination, more concentrated than any other sets. This shot pattern is a very typical and effective pattern in tennis. Further analysis shows that this pattern was used mostly by KA, with good results. It is interesting that this three-shot only appeared twice in the third and fourth sets. Perhaps PCB made necessary adjustments so this tactic is less effective for KA. Another interesting pattern is that there are more forehand returns than backhand returns in the two-shot, three-shot, and four-shot space for the first set, which PCB won. In the second, third, and fourth sets that KA won, there are significantly more backhand returns than forehand returns in the three-shot and four-shot space. In the two-shot space, the forehand and backhand returns were roughly equal for the second, third, and fourth sets. Is there a correlation between KA's winning the second, third, and fourth sets and more backhand returns? Did KA adjust his serves to create such patterns? It is an interesting questions for further analysis.

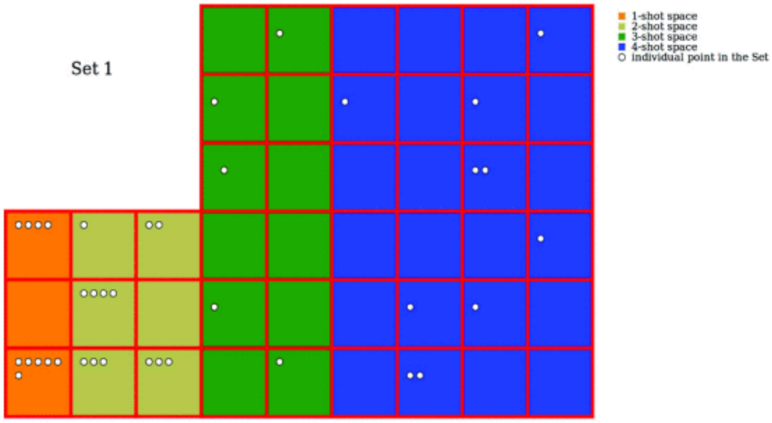

Fig. 3. Points Plotted for the First set.

C. Service Games Analysis

The fractal tables in figures 7 and 8 show the first and second serves for both players. From the first serve table (Figure 7), we can see the contrast between the two players. KA had more powerful serves because he had far more aces than PCB in the one-shot space. However, PCB had advantage in the two-shot space. This means that although PCB's serves were not as fast as KA's, he was still able to force KA into making many return errors. In addition, KA never served to the body and slightly preferred serving to the T. KA's serves were more predictable but due to his power, he was still able to serve many aces. On the other hand, PCB's serves were more evenly distributed among the three different directions. It seems that PCB was trying to make his serves more unpredictable. In the four-shot space, we see far more PCB's first serves went to AK's backhand side than AK's forehand side. While in the three-shot space, it is evenly distributed. It is not immediately clear whether this pattern is tactically significant, but it is an interesting question to explore. In the three-shot and four-shot spaces, we can see several clusters of points. These seem to be the shot patterns that both players favor. For example, in the three-shot space, we have the frequent patterns WBF, WFF, and TFF, particularly WBF. In the four-shot space, we have the frequent patterns TBFF and BBFB. From a coaching perspective, the player may need to practice these patterns more than others in training. From the second serve table (Figure 8), we see that both players serve far more to the opponent's backhand than to the forehand. Again, KA frequently used the three-shot pattern WBF with his second serves, while PCB did not use this pattern at all.

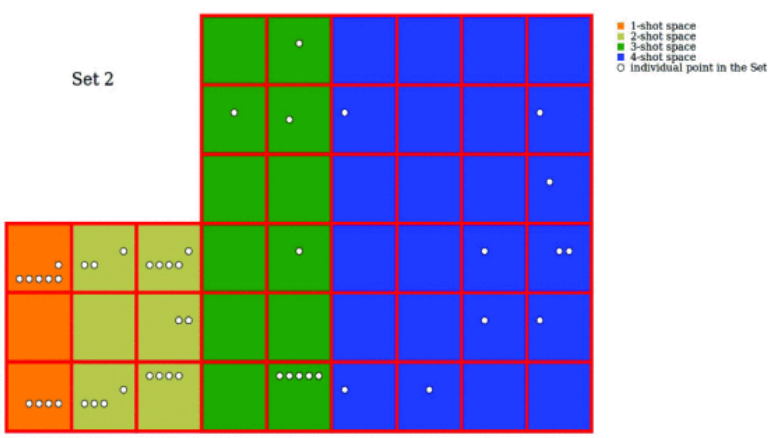

Fig. 4. Points Plotted for the Second set.

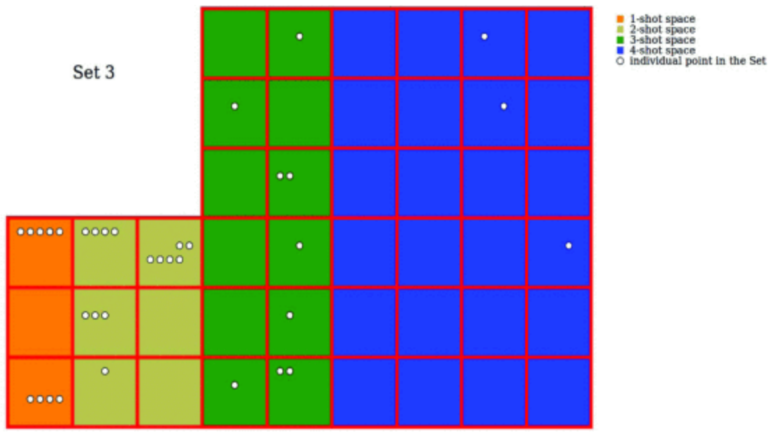

Fig. 5. Points Plotted for the Third set.

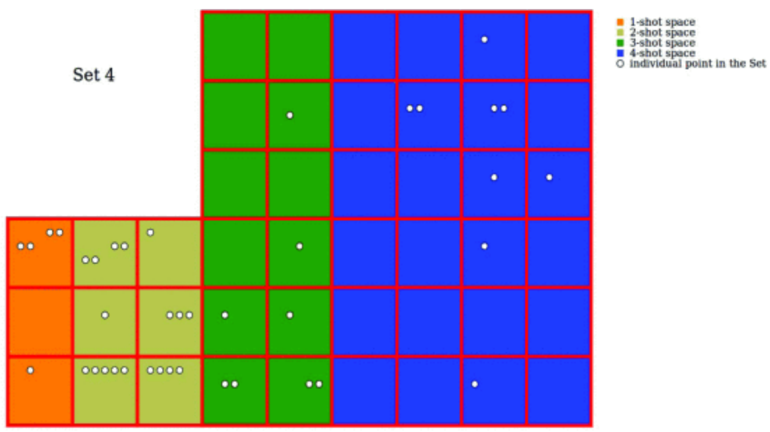

Fig. 6. Points Plotted for the Fourth set.

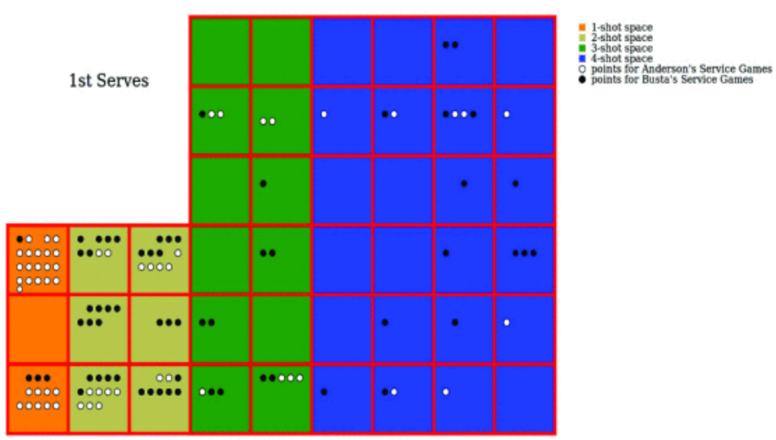

Fig. 7. First Serve Service Games for both players.

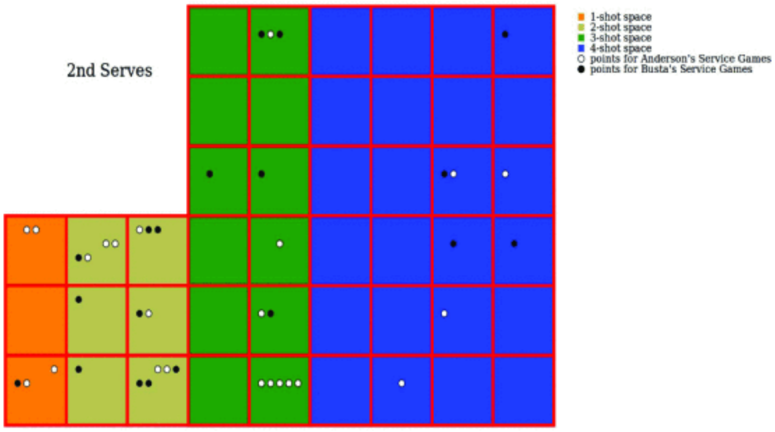

Fig. 8. Second Serve Service Games for both players.

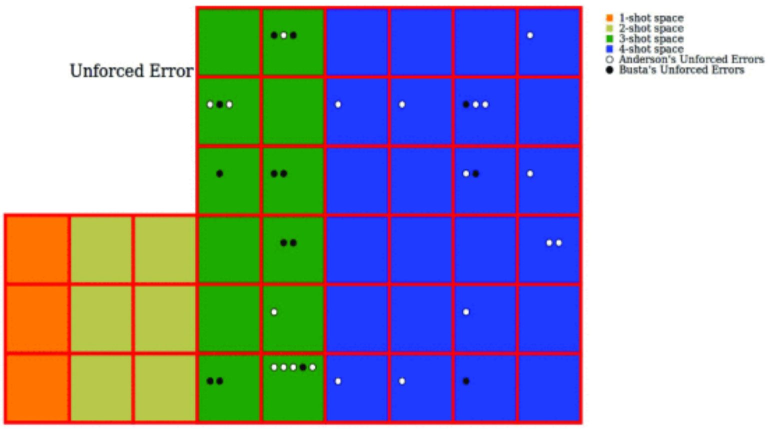

Fig. 9. Unforced Errors for both players.

D. Point Outcome Analysis

Figure 9 shows the unforced errors in the fractal table. Unforced error is one of the most important topics of tennis performance analysis because it is an indicator of a player's form. It is also something a player may be able to control. From the data visualization, we can see an interesting pattern: PCB made far more unforced errors in the three-shot space than in the four-shot space, while KA made far more unforced errors in the four-shot space than the three-shot space. Does it suggest that PCB is better at playing longer points? In the four-shot space, PCB's few unforced errors all involved forehand preceded by a backhand return. Were the backhand returns the cause of the unforced errors in his forehand? KA's unforced errors in the four-shot space seem more scattered but four unforced errors were related to a forehand in the third shot.

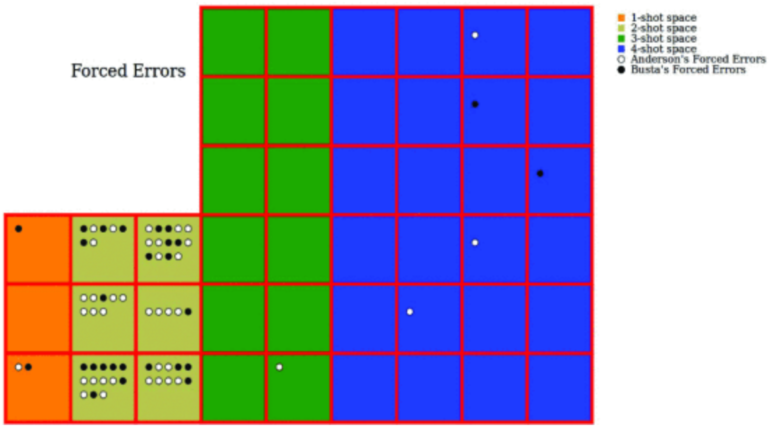

Fig. 10. Forced Errors for both players.

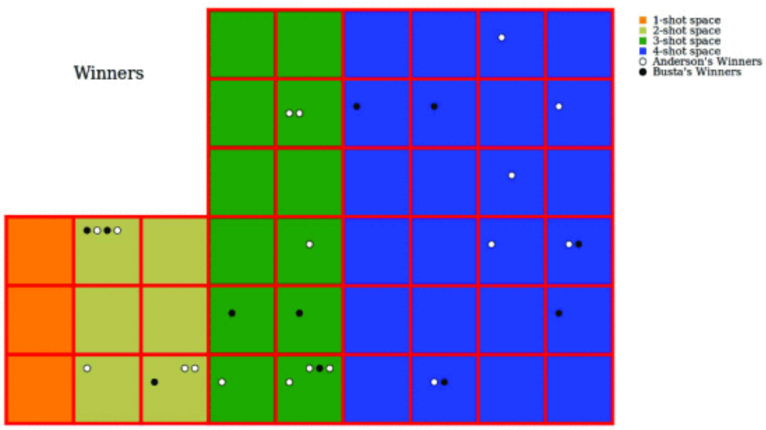

Fig. 11. Winners scored by both players.

In Figure 10, we see that most forced errors happen in the two-shot space, most likely the result of both player's aggressive serves. In the two-shot space, KA has far more forced errors than PCB when returning a serve to the body. This is likely a weakness of KA, perhaps because he is taller and less nimble than PCB.

The fractal table in figure 11 shows the winners by both the players. In the three-shot and four-shot spaces, KA's winners all involved a backhand return. PCB's winners in the three-shot space all involved serving wide.

E. Player's Net Profit or Loss Analysis

We want to study which player has advantage in which shot patterns. To do this, we calculate the win-loss difference between the two players in each shot pattern box and call it net profit or net loss for each box. We use red color to denote a net loss and use black color to denote a net profit. White color means break-even. We use the opacity of the red or black color to encode the differences. For example, a bright red color means a big loss while a light red color means a small loss. The fractal tables in Figures 12 and 13 show the net profit or net loss for KA and PCB respectively. In Figure 12 and 13, we can see that KA served far more aces than PCB, particularly to the T. However, in the two-shot space, PCB had advantage. This suggests that PCB's return of serves were better than KA's. In the three-shot space, KA has a slight advantage, suggesting that KA's third sot (i.e. the shot after the serve and the return) is slightly more effective than PCB's, perhaps due to KA's powerful serves. In the four-shot space, PCB has a small advantage. Overall, the difference between KA and PCB's performance after the serve is quite small. This suggests that neither player has a significant advantage in particular patterns, except for the serves, which KA had clear advantage. The serves seem to be the deciding factor for KA to win this match. On the other hand, the profit/loss patterns seem to suggest that PCB had advantage when more shots were played. Therefore, a strategy can be derived from our visualizations: PCB should focus on returning KA's serve well and try to player longer points. On the other hand, the data visualizations suggest that KA should focus on serving well (particularly to the T) and try to avoid unforced errors in longer rallies (e.g. four shots or higher).

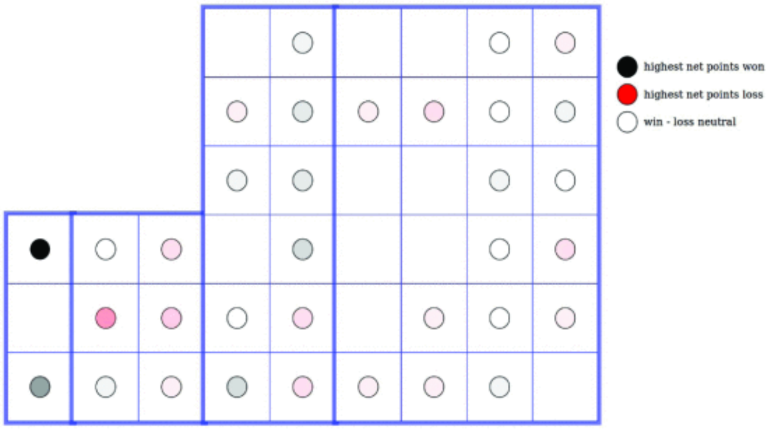

Fig. 12. Anderson's Net Profit or Loss.

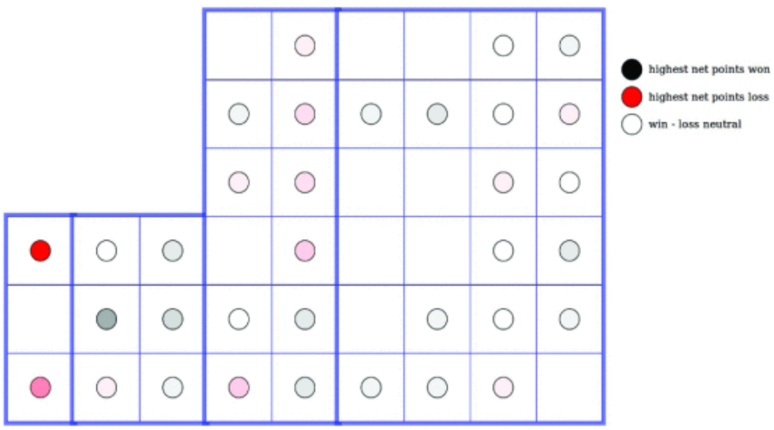

Fig. 13. Busta's Net Profit or Loss.

SECTION VI. CONCLUSIONS

In this paper we described a method to analyze and visualize the shot patterns of tennis matches. We have developed a novel fractal table to visualize the tennis shot patterns and their match statistics. Compared with the conventional high-level tennis match statistics, our method provides a micro-level analysis that reveals the complicated dynamics of a tennis match. As shown in our case studies, many interesting patterns can be found through the fractal table data visualization. Such data visualizations may not provide the answers, but they are useful for generating intriguing questions that lead to further analyses of the players' strengths and weakness. This tool can be used by tennis experts, players, coaches and serious fans to analyze and compare a player's performance. In the future, we plan to expand this micro-level analysis to other aspects of tennis matches and apply this approach to other sports if such micro-level data is available.

References

1. X. He and Y. Zhu, "TennisMatchViz: a tennis match visualization", *Proceedings of International Conference on Visualization and Data Analysis (VDA)*, 2016.

2. S. Pokharel and Y. Zhu, "Analysis and Visualization of Sports Performance Anxiety in Tennis Matches", *Proceedings of the 13th International Symposium on Visual Computing Lecture Notes in Computer Science*, vol. 11241, pp. 407-419, 2018.
3. M. Madurska, "A set-by-set analysis method for predicting the outcome of professional singles tennis matches", *Masters Thesis*, 2012.
4. A. J. O'Malley, "Probability Formulas and Statistical Analysis in Tennis" in Journal of Quantitative Analysis in Sports, De Gruyter, vol. 4, pp. 1-23, 2008.
5. P. Newton and J. Keller, "Probability of Winning at Tennis I", *Studies in Applied Mathematics Theory and data Studies in Applied Mathematics*, vol. 114, pp. 241-269, 2005.
6. L. Riddle, "Probability Models for Tennis Scoring Systems", *Journal of the Royal Statistical Society.*, vol. 37, pp. 63-75, 1988.
7. D. Jackson and K. Mosurski, "Heavy defeats in tennis: Psychological momentum or random effect?", *CHANCE*, vol. 10, pp. 27-34, 1997.
8. I. MacPhee, J. Rougier and G. Pollard, "Server advantage in tennis matches", *Journal of Applied Probability*, vol. 41, pp. 1182-1186, Dec 2004.
9. T. Polk, J. Yang, Y. Hu and Y. Zhao, "TenniVis: Visualization for Tennis Match Analysis", *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, pp. 2339-2348, 2014.
10. J. Liqun and D. G. Banks, "TennisViewer: a browser for competition trees", *IEEE Computer Graphics and Applications*, vol. 17, pp. 63-65, 1997.
11. M. Burch and D. Weiskopf, "Tennis Plots: Game Set and Match" in Diagrammatic Representation and Inference Lecture Notes in Computer Science, Springer, vol. 8578, pp. 38-44, 2014.
12. X. Wei, P. Lucey, S. Morgan, P. Carr, M. Reid and S. Sridharan Machar, "Predicting Serves in Tennis Using Style Priors", *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2207-2215, 2015.
13. X. Wei, P. Lucey, S. Morgan and S. Sridharan, "Forecasting the Next Shot Location in Tennis Using Fine-Grained Spatiotemporal Tracking Data", *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, pp. 2988-2997, 2016.
14. Y. Wu, J. Lan, X. Shu, C. Ji, K. Zhao, J. Wang, et al., "iTTVis: Interactive Visualization of Table Tennis Data", *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, pp. 709-718, 2018.
15. Y. Wu, X. Xie, J. Wang, D. Deng, H. Liang, H. Zhang, et al., "ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer", *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 65-75, 2019.
16. [online] Available: https://github.com/JeffSackmann/tennis_ MatchChartingProject.
17. ATP Stats, [online] Available: https://www.atpworldtour. com/en/stats.
18. "Design Driven Documents", [online] Available: https://d3js.org/.
19. H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim and O. Deussen, "Feature-Driven Visual Analytics of Soccer Data", *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology(VAST)*, pp. 13-22, 2014.