Scanpath Comparison Revisited

Andrew T. Duchowski[⋆], Jason Driver[⋆], Sheriff Jolaoso[†], Beverly N. Ramey[‡], Ami Robbins[‡], and William Tan[∗]

[⋆]School of Computing, Clemson University; [†]Computer Engineering, UMBC; [‡]Computer Science, Winthrop University; and [∗]Computer Engineering, Syracuse University

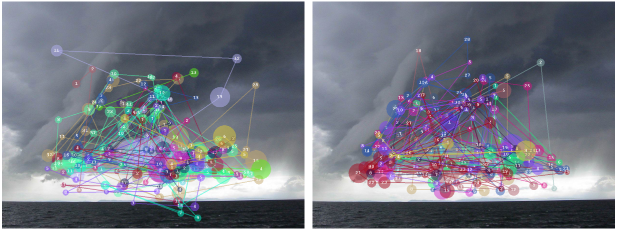
ETRA 2010, 22-24 March, Austin, TX

Abstract

Abstract

The scanpath comparison framework based on string editing is revisited. The previous method of clustering based on *k*-means "preevaluation" is replaced by the mean shift algorithm followed by elliptical modeling via Principal Components Analysis. Ellipse intersection determines cluster overlap, with fast nearest-neighbor search provided by the *k*d-tree. Subsequent construction of *Y*-matrices and parsing diagrams is fully automated, obviating prior interactive steps. Empirical validation is performed via analysis of eye movements collected during a variant of the Trail Making Test, where participants were asked to visually connect alphanumeric targets (letters and numbers). The observed repetitive position similarity index matches previously published results, providing ongoing support for the scanpath theory (at least in this situation). Task dependence of eye movements may be indicated by the global position index, which differs considerably from past results based on free viewing.

Introduction



- Scanpaths are compelling visualizations of eye movements (Noton & Stark, 1971)
- Not yet fully exploited for quantitative potential
- Want an easy computation analogous to ANOVA table
- I like string editing approach for computing “similarity” between scanpath pairs (Privitera & Stark, 2000)
- Resulting metric is similar to Spearman’s rank-order coefficient (Boslaugh & Watters, 2008) but with coefficient $S \in [0, 1]$ instead of $S \in [-1, 1]$

Background: What's been done before?

- Scanpaths have been compared to evaluate on-screen television enhancements (Josephson & Holmes, 2002, 2006)
- String editing used to investigate amalgamation of scanpaths into a single, representative scanpath (Hembrooke et al., 2006)
- Levenshtein similarity replaced by Needleman-Wunsch distance yielding *eyePatterns* (West et al., 2006)
  - note that "large distance" is "small similarity" (Waterman, 1989)
- Other approaches are limited in functionality (e.g., *ProtoMatch* (Myers & Schoelles, 2005)) or use trajectory-based approach (Vlachos et al., 2002, 2004; Torstling, 2007)

Comparison with recent work

- Our contribution is similar to Jarodzka et al. (2010)...
  - mean-shift clustering similar to amplitude-based clustering
  - pairwise Y-matrix similar to adjacency matrix $A(k, l)$
- ...but with important differences
  - no use of shortest path (Dijkstra's) although that's an interesting approach
  - not based on either *gridded* or *semantic* AOIs ("content-driven") but rather clustered on fixation points themselves ("data-driven")
- Other important (but perhaps subtle) points to remember
  - temporal information is taken into account (via temporal component of mean-shift cluster)
  - pairwise statistical significance is calculated against random scanpaths—significance between scanpath groups not yet implemented (but I don't think at all intractable)

Example of Approach

- String editing was used by Privitera and Stark (2000) to compare human fixations with those predicted by automatic means (e.g., Itti et al.'s (1998) saliency model)

$s_1 = abcfeffgdc$ $s_2 = afbffdcdf$	`start`	`cost 0`
$s_1 = abcfeffgdc$ $s_2 = afeffdcdf$	`after substitution of first` b `by` e	`cost 1`
$s_1 = abcfeffgdc$ $s_2 = abcfeffdcdf$	`after insertion of` bc `after first` a	`cost 2`
$s_1 = abcfeffgdc$ $s_2 = abcfeffdc$	`after deletion of last` df	`cost 2`
$s_1 = abcfeffgdc$ $s_2 = abcfeffgdc$	`after insertion of` g	`cost 1`

Implementation: Levenshtein ~~Distance~~ Similarity

- Given two strings $s_1 = abcfeffgdc$ and $s_2 = afbffdcdf$, construct 10×9 array A
- Assign cost of character *deletion*, *insertion*, or *substitution*:

$$A[i][j] = \min \begin{cases} A[i-1][j] + 1 & \text{// deletion} \\ A[i][j-1] + 1 & \text{// insertion} \\ A[i-1][j-1] + c(i,j) & \text{// substitution} \end{cases}$$

where

$$c(i,j) = \begin{cases} 0, & s_1[i-1] = s_2[j-1] \\ 1, & \text{otherwise} \end{cases}$$

- Note that costs can be weighted, as in Needleman-Wunsch distance

Implementation: Dynamic Programming

- Use dynamic programming to arrive at transformation cost
- Normalize total cost to the length of the longer string, in this case 9, yielding $S_s = (1 - 6/9) = 0.33$
- Store pair-wise similarity coefficients in a table, the Y-matrix, with one row and column for each image viewed by a subject (for all subjects)

	a	f	b	f	f	d	c	d	f
a	0	1	2	3	4	5	6	7	8
b	1	1	1	2	3	4	5	6	7
c	2	2	2	2	3	4	4	5	6
f	3	2	3	2	2	3	4	5	5
e	4	3	3	3	3	3	4	5	6
f	5	4	4	3	3	4	4	5	5
f	6	5	5	4	3	4	5	5	5
g	7	6	6	5	4	4	5	6	6
d	8	7	7	6	5	4	5	5	6
c	9	8	8	7	6	5	4	5	**6**

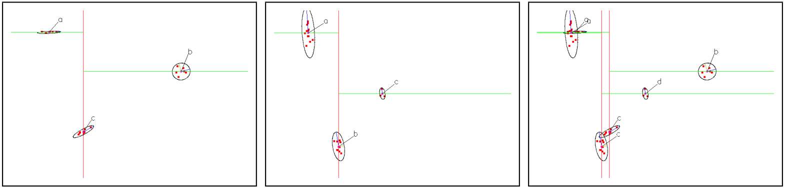
Implementation: Data Organization

	Subj. 1 Pict 1	Subj. 1 Pict 2	Subj. 2 Pict 1	Subj. 2 Pict 2
S1P1	**R**	**I**	**L**	**G**
S1P2		**R**	**G**	**L**
S2P1			**R**	**I**
S2P2				**R**

	Same Subj. (SS)	Diff. Subj. (DS)
Same Img. (SI)→	**R**epetitive	**L**ocal
Diff. Img. (DI)→	**I**diosyncratic	**G**lobal
		Random

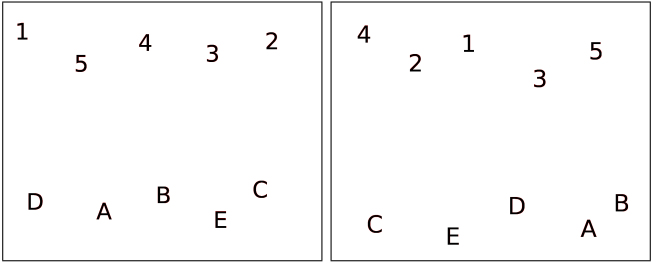
- String editing used to quantitatively measure loci of fixations S_p as well as order S_s
- Similarity coefficients stored in Y-matrix
- Values from Y-matrix condensed (averaged) in two tables, called Parsing Diagrams
- Two parsing diagrams, one for each of S_p and S_s indices

Innovations: Labeling & Fast Overlap Lookups



- Improve scanpath comparison by substituting *k*-means clustering with mean shift (Santella & DeCarlo, 2004)
  - *k* means requires *a priori* knowledge of the number of clusters (Duda & Hart, 1973)
  - mean shift is “self-organizing” in comparison
- Use Principal Components Analysis to model elliptical cluster boundaries
  - use ellipses to calculate overlap among clusters
  - use *k*d-tree to spatially partition scanpath clusters for efficient nearest-neighbor search (Hoppe, 1994)

Empirical Validation



- An experimental paradigm was sought to elicit similar scanpaths from participants
- A gaze-directed variant of the Trail Making Test protocol (Bowie & Harvey, 2006) was chosen
- The TMT is usually comprised of parts A and B
  - part A: 1-2-3-4-5-A-B-C-D-E
  - part B: 1-A-2-B-3-C-4-D-5-E

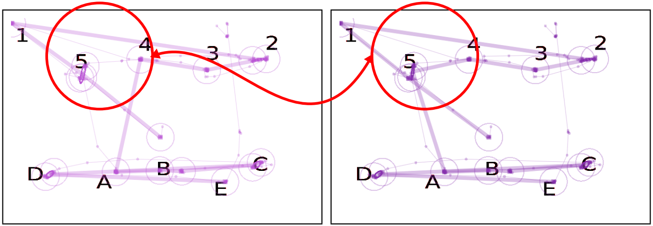
More on the TMT

- The TMT is thought to measure processing speed, sequencing, mental flexibility, and visual-motor skills
  - Part A is presumed to be a test of visual search and motor speed skills
  - Part B is considered to also test higher level cognitive skills
- Normally, the TMT's main dependent variable of interest is total time to completion
- In its present instantiation, the primary measure of interest is the scanpath (which inherently encodes processing time)
- Main concerns here are **spatial distribution** and **ordering**
- Repetitive scores are obtained by recording two scanpaths over a single image
- Local and global indices are gathered by having multiple participants perform the test

Experimental Design

- Subjects: six college students (4 M, 2 F; ages 18-27, median age 21)
  - results from the TMT should be stratified by age and education (Tombaugh, 2004); our sample represents one such strata
- Stimulus: two 1280 $\times$ 1024 images
- Procedure: 5-point calibration sequence, followed by TMT-A, and TMT-B, each image viewed twice (order not counterbalanced)
  - participants were asked to view the sequences as quickly as possible but dwelling over each number or letter for a fraction of a second (they were aware of the underlying fixation algorithm)
- Apparatus: Tobii ET-1750 video-based corneal reflection (binocular) eye tracker

Pilot Testing



- Mean shift clustering of fixations $\mathbf{x}_i = (x_i, y_i, t_i)$ depends on the use of a kernel function (Santella & DeCarlo, 2004)

$$K([\mathbf{x}_i, t_i]) = \exp\left(\frac{x_i^2 + y_i^2}{\sigma_s^2} + \frac{t_i^2}{\sigma_t^2}\right)$$

  where σ_s and σ_t determine local support of the kernel in both spatial (dispersion) and temporal extent
- Pilot testing revealed the importance of both spatial and temporal support

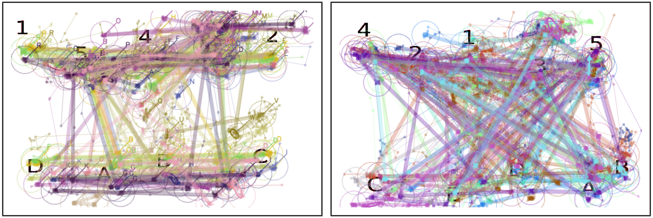
Aggregate Results

	SS	DS
SI →	**R**epetitive 0.65 F(1,22) = 98.2, $p < 0.01$	**L**ocal 0.47 F(1,238) = 848.2, $p < 0.01$
DI →	**I**diosyncratic 0.44 F(1,46) = 165.4, $p < 0.01$	**G**lobal 0.44 F(1,238) = 884.0, $p < 0.01$
	S_p	Random 0.06

	SS	DS
SI →	**R**epetitive 0.35 F(1,22) = 34.6, $p < 0.01$	**L**ocal 0.23 F(1,238) = 148.5, $p < 0.01$
DI →	**I**diosyncratic 0.18 F(1,46) = 52.1, $p < 0.01$	**G**lobal 0.17 F(1,238) = 221.0, $p < 0.01$
	S_s	Random 0.08

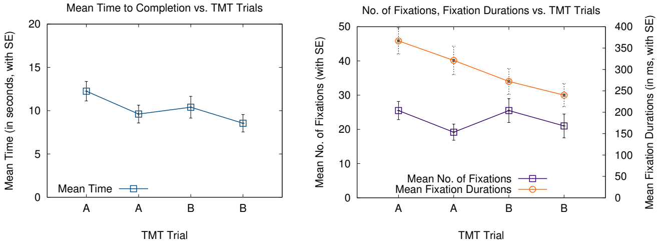
- Statistical significance derived from random scanpath comparisons
- Position indices > sequence indices
- Repetitive indices show highest correlations
- Repetitive position index is comparable to previous work (0.65 vs. 0.64)
- Key difference here is task (TMT vs. free viewing)
- Global position index may indicate task dependence

Segregate Results: TMT-A vs. TMT-B



- Aggregate statistics tend to obscure processes related to individual behaviors or stimuli
- Analysis over just TMT-A and TMT-B shows that repetitive (and local) scores are higher for TMT-A
- TMT-A relies mainly on visual search and should therefore be easier to execute (fewer errant saccades)

Perf. & Process Metrics Across Trials



- TMT-B's lower S_p, S_s suggest increased cognitive difficulty
- However, repeated measures ANOVA only shows a marginally significant main effect of trial on speed, and . . .
- . . . time to completion decreases, suggesting decreased cognitive difficulty (opposite of what was expected)
- Process measures suggest **learning effect** as fixation durations decrease significantly across trials but the number of fixations do not

Study Notes & Limitations

- Aggregate analysis of the six participants' scanpaths shows position indices are generally higher than sequence indices, as expected
- Repetitive indices show highest correlations (not surprising given the task stipulated by the Trail Making Test protocol)
- A shortcoming of the framework is lack of significance testing between different groups of scanpaths
  - Feusner and Lukoff (2008) suggest computation of the $d^* = d_{between} - d_{within}$ statistic, where $d_{between}$ is the average distance between scanpaths in different groups and d_{within} is the average distance between scanpaths in the same group
  - to include this computation within present framework would require construction of additional between-group and within-group Y-matrices

General Discussion

- Scanpath comparison adds another dimension to traditional speed/performance analysis
- Quantification of position and order similarity appears to provide useful information (e.g., pointing out similarity to random order)
- Aggregate analysis may be prone to "saturation effect" (e.g., given too many scanpaths, numbers may become meaningless—entire area will be covered visually eventually)
- Segregate analysis may be more meaningful (e.g., between-subjects analysis for different cultural groups or expert/novice comparisons)

Conclusion

- Mean-shift clustering of fixations and elliptical modeling enables automation of the string editing approach
- Construction of a kd-tree facilitates efficient lookup ($O(\log n)$ average time per search)
- The combination of these algorithms removes prior reliance on preevaluation and human intervention
- Scanpath comparison metrics validated empirically by a variant of the Trail Making Test
- For a well-defined visual task, moderately correlated repetitive and global position indices are expected

Acknowledgments

This work was supported in part by CNS grant #0850695 from the National Science Foundation (REU Site: Undergraduate Research in Human-Centered Computing).

Questions

- Thank you
- Comments, Questions?

Selected References I

Boslaugh, S., & Watters, P. A. (2008). *Statistics in a Nutshell*. Sebastopol, CA: O'Reilly Media, Inc.

Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the Trail Making Test. *Nature Protocols*, *1*(5), 2277–2281.

Duda, R. O., & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York, NY: John Wiley & Sons, Inc.

Feusner, M., & Lukoff, B. (2008). Testing for statistically significant differences between groups of scan patterns. In *Eye Tracking Research & Applications (ETRA) Symposium* (pp. 43–46). New York, NY.

Selected References II

Hembrooke, H., Feusner, M., & Gay, G. (2006). Averaging Scan Patterns and What They Can Tell Us. In *Eye Tracking Research & Applications (ETRA) Symposium* (p. 41). San Diego, CA.

Hoppe, H. (1994). *Surface Reconstruction From Unorganized Points*. Unpublished doctoral dissertation, University of Washington, Seattle, WA.

Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, *20*(11), 1254-1259.

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A Vector-based, Multidimensional Scanpath Similarity Measure. In *Eye Tracking Research & Applications (ETRA) Symposium* (p. 211-218). Austin, TX.

Selected References III

Josephson, S., & Holmes, M. E. (2002). Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web. In *Eye Tracking Research & Applications (ETRA) Symposium* (p. 43-49). New Orleans, LA.

Josephson, S., & Holmes, M. E. (2006). Clutter or Content? How On-Screen Enhancements Affect How TV Viewers Scan and What They Learn. In *Eye Tracking Research & Applications (ETRA) Symposium* (p. 155-162). San Diego, CA.

Myers, C. W., & Schoelles, M. J. (2005). ProtoMatch: A tool for analyzing high-density, sequential eye gaze and cursor protocols. *Behavior Research Methods, Instruments, Computers (BRMIC)*, *37*(2), 256-270.

Selected References IV

Noton, D., & Stark, L. (1971). Scanpaths in Saccadic Eye Movements While Viewing and Recognizing Patterns. *Vision Research*, *11*, 929-942.

Privitera, C. M., & Stark, L. W. (2000). Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, *22*(9), 970-982.

Santella, A., & DeCarlo, D. (2004). Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest. In *Eye Tracking Research & Applications (ETRA) Symposium* (p. 27-34). San Antonio, TX.

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*, 203–214.

Selected References V

Torstling, A. (2007). *The Mean Gaze Path: Information Reduction and Non-Intrusive Attention Detection for Eye Tracking*. Unpublished master's thesis, The Royal Institute of Technology, Stockholm, Sweden. (Techreport XR-EE-SB 2007:008)

Vlachos, M., Gunopulos, D., & Das, G. (2004). Rotation invariant distance measures for trajectories. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 707–712). New York, NY.

Vlachos, M., Kollios, G., & Gunopulos, D. (2002). Discovering Similar Multidimensional Trajectories. In *ICDE '02: Proceedings of the 18th International Conference on Data Engineering* (pp. 673–685). Washington, DC.

Selected References VI

Waterman, M. S. (1989). Sequence Alignments. In M. S. Waterman (Ed.), *Mathematical Methods for DNA Sequences* (pp. 53–92). Boca Raton, FL: CRC Press, Inc.

West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). eyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences. In *Eye Tracking Research & Applications (ETRA) Symposium* (p. 149-154). San Diego, CA.