Recognizing Gaze-Motor Behavioral Patterns in Manual Grinding Tasks

Gregory Bales[1], Jayanti Das[1], Barbara Linke[1], and Zhaodan Kong[1*]

[1]The Department of Mechanical and Aerospace Engineering, University of California, Davis, Davis, CA, USA
glbales, jydas, bslinke, zdkong@ucdavis.edu

Abstract
This paper reports our progress in developing techniques for "parsing" raw gaze and force data from manual grinding tasks into a principled model. A grinding task, though simple, requires the practitioner to combine elements from the large repertoire of her skillset. Based on the joint, gaze, and force data collected from a series of experiments, and by extending existing scanpath methods, we develop a visualization method called Gaze-Motor Space-Time Cube (GMSTC), which can help us gain insight into the joint gaze-motor routine existing in complex manual tasks. For instance, there exists a strong correlation between the spectra of a subject's fixation and force distributions. Such insight might be hard to extract through an examination of either the gaze or the force data separately. Furthermore, by comparing data obtained from operators with different levels of skill, we are able to quantitatively describe characteristics of human manual skill. For instance, we find that an experienced subject exhibits longer fixation durations and smaller fixation variations than an intermediate one. A detailed understanding of gaze-motor behavior broadens our knowledge of how a manual task is executed. Our results help to provide this extra insight, and have implications in the way in which knowledge and manual expertise is transferred from one generation of practitioners to the next.

Keywords: Human factors, man-machine interactions, data visualization, attention, grinding

1 Introduction

This paper studies human manual skills involved in manufacturing tasks. Despite the current proliferation of highly automated factories, humans are still an integral part of most manufacturing processes, not only as the supervisors of machines, but also as the performers of manual tasks, such as fine polishing. The skills involved in these manual tasks are largely procedural rather than declarative, meaning that they cannot be easily articulated by the individuals (Goldstein, 2014). A lack of understanding of these manual skills may prolong the transfer of this knowledge from generation to generation. It may also impede the development of novel manufacturing themes, such as cloud

[*] Corresponding author.

manufacturing (Lee et al., 2015) and human-robot collaborations (Wojtara et al., 2009). To elaborate on the last point, in order to establish an effective dialog within a human-robot team, the robot must be able to understand the intention of the human by inferring from her behavior. In this paper, we focuses on analyzing the visual-attention-motor behavior in the context of manual grinding tasks (Odum et al., 2014), (Klocke et al., 2011). Our study, on the one hand, improves our understanding of complex manual skills that are beyond our everyday activities and, on the other hand, enables models of human skills which are crucial for building future industrial robots that can understand human intentions.

Visual attention is a remarkable human capability of reducing the huge amount of visual data entering our eyes into a manageable level. Given the nonuniform distribution of photoreceptors in our retina (with a high-resolution central fovea and a low-resolution periphery), the vision system cannot reconstruct a detailed 3D model of the scene. Instead, visual attention guides our eyes to important parts of the scene and puts the object of interests in the fovea (Borji and Itti, 2013). In the past few decades, a considerable amount of experimental and computational research has been conducted to understand and model visual attention process (Borji and Itti, 2013), (Rayner, 2009). Aside from being an interesting scientific problem itself, visual attention also attracts significant interest from engineering fields especially but not limited to robotics (Begum and Karray, 2011), (Siagian and Itti, 2009), computer vision (Guo and Zhang, 2010), and human-robot collaborations (Bauer et al., 2008), (Wojtara et al., 2009).

Visual attention can be roughly divided into two categories, bottom-up attention and top-down attention (Borji and Itti, 2013), (Posner, 2011). Top-down or voluntary attention is our ability to intentionally attend to something. It is a goal-driven process based on aspects such as tasks, knowledge, expectations, and memory. In a manual grinding scenario, for instance, it may involve finding a sample. Bottom-up or reflexive attention is a stimulus-driven process in which a salient sensory event captures our attention. Such an event might be a crack appearing on the surface of the grinding sample. A popular bottom-up attention model is the saliency map model, based on intrinsic image features such as color, orientation, and intensity (Itti et al., 1998), (Borji and Itti, 2013). However, it has been argued and demonstrated that at least with purposeful actions, top-down attention rather than bottom-up attention plays a much dominant role (Land, 2009), (Hayhoe and Ballard, 2005), (Borji et al., 2014). Some interesting observations from the studies of top-down attention in natural behavior, such as making a cup of tea (Land, 2009) and making a sandwich (Borji et al., 2014), include: the first movement of the hand toward the object of interest occur before the first saccade; at the end of each object-related action, the eyes move to the next object before the manipulation is over. In this paper, we focus on top-down attention involved in manual manufacturing tasks.

Gaze, a coordinated motion of the eyes and the head, is often used in research as a proxy for visual attention[†]. The majority of studies are concerned with two types of gaze movements, saccades, fast ballistic movements in which the gaze moves rapidly from one location to another, and fixations, the period during which the gaze is stationary and useful information is collected (Duchowski, 2007), (Cristino et al., 2010). Sequences of time-ordered gaze movements composed of fixations and saccades, called scanpaths, have been used for visualizing and analyzing gaze since the early 1970s (Noton and Stark, 1971), (Duchowski et al., 2010). Scanpaths facilitate the quantitative comparison of gaze data collected from different subjects and trials. One algorithm that can be used for such comparisons is based on a metric called Levenshtein distance, more commonly referred to as string editing distance. The string edit algorithm (SEA) compare scanpaths by solving an optimization problem with unit costs assigned to three different character operations: deletion, insertion, and substitution (Privitera and Stark, 2000), (Duchowski et al., 2010). Some issues with SEA are that it cannot differentiate between close and distance regions and it does not take into account fixation

[†] It should be pointed out that gaze, which can be tracked by an eye-tracker, corresponds to the overt movements of the eyes, not the covert movements of visual attention. Thus, a very important assumption that is usually accepted in visual attention research is that the attention is linked to the gaze direction, even though this might not be always true (Duchowski, 2007).

duration. To solve these issues, the ScanMatch method encodes both temporal and spatial information into its representation of gaze sequences and compares two scanpaths by maximizing the similarity score computed from a substitution matrix that assigns positive values for matches and negative values for mismatches (Cristino et al., 2010). Other scanpath comparison methods developed over the years include sequential pattern mining (SPAM) algorithm (Hejmady and Narayanan, 2012), dot-plot method (Goldberg and Helfman, 2010), binomial tests of Markov matrices (Underwood et al., 2003), vector-based approach (Jarodzka et al., 2010), and maximum transition based agglometrative hierarchical clustering (MTAHC) algorithm (Kang and Landry, 2015).

It is widely accepted that visual attention is not decoupled from motor system in natural behavior (Land, 2009), (Erez et al., 2011), (Yi and Ballard, 2009), (Sprague et al., 2007). Taking hand-eye coordination as an example, on the one hand, the eyes' saccades disambiguate the scene in a task-relevant manner and, on the other hand, the hands' motions anticipate the eyes' saccades (Erez et al., 2011). In the majority of studies concerned with visual attention and motor system, actions are discrete, e.g., "remove the lid of the kettle" and "select a peanut butter jar" (Land, 2009), (Yi and Ballard, 2009), and manually labeled by humans. Such a representation fails to capture the complex nature of gaze-motor behavior. In these studies, only the onsets of actions were considered. Data on motor dynamics, such as the changes in forces, were not collected and subsequently studied. A representation that captures the dynamic nature of motor behavior is needed, similar to those developed in (Kong and Mettler, 2013), (Mettler et al., 2014). This paper presents a new method that visualizes data collected on the joint gaze and motor behavior and compares the behaviors of individuals with different skill levels.

The contribution of the paper is two-fold. First of all, we propose a new visualization method that can help us to gain insight into the joint gaze-motor routine existing in complex manual tasks. Second, we quantitatively characterize manual skills by comparing joint gaze-motor data. To our knowledge, this paper is one of the first instances in which visual attention has been studied in manufacturing scenarios. Adept hand eye coordination is key to the performance of a number of manufacturing processes. Due to the importance of gaze-motor behavior, our method can be generalized to gain insight into a wide range of industrial activities such as welding, repairing machinery, grinding and polishing during abrasive finishing process, or everyday activities like driving.

2 Experiment

In this section, we describe the setup and procedure of our manual grinding experiment.

2.1 Setup

For the purpose of studying manual skills involved in manufacturing tasks, we recruited two students from the Department of Mechanical and Aerospace Engineering, University of California, Davis. The two subjects were chosen based on their differing levels of experience. For this study we have subjectively defined experience as the amount of time each subject has spent with grinding tools. In the subsequent sections, the "experienced" subject shall be referred to as Subject 1, and the "intermediate" subject as Subject 2. Each subject was asked to use an abrasive wheel to grind a metal sample to a smooth finish to the best of their ability. Three trails were performed by each subject. As shown in Figure 1 two streams of data were collected. First, we measured the direction of gaze with an eye tracking system. Second, the grinding forces were measured with a triaxial load cell mounted beneath the grinding sample.

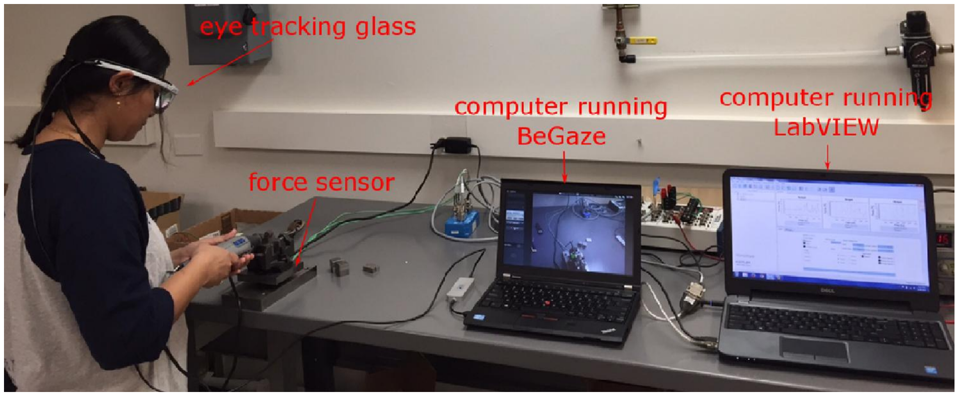

Figure 1: Setup of our grinding experiment. The figure shows one of our tested subjects using an abrasive wheel to grind a metal sample. While a person was using an abrasive wheel to grind a metal sample, data were collected from two modules: 1) a gaze tracking module, consisting a pair of SMI eye-tracking glasses and a computer running BeGaze algorithm, 2) a force measuring module, consisting of a force sensor and a computer running LabVIEW. The data collected from the two modules were synchronized and then analyzed using the methods described in Sections 3,4, and 5.

We measured gaze using a wearable eye tracking system manufactured by SensoMotoric Instruments (SMI). The SMI ETG 2w system is integrated into a set of glasses which can extract binocular gaze while simultaneously recording a video of the visual point of view. Pupil images and corneal reflection points are used to determine the vertical and horizontal angular orientation of each individual eye, which in turn are used to calculate the gaze. True gaze direction requires a vector to describe its full nature. In our analysis, the gaze data were represented as a binocular points of regard (BPOR). These points describe where the binocular gaze vector pierced the gaze plane, a hypothetical projection plane located 1450mm in front of the glasses. The BPOR was sampled at 60 Hz and the gaze was presented as their pixel positions within the video image. The video resolution was 1280 pixels horizontally and 960 pixels vertically, and was recorded at 60 frames per second.

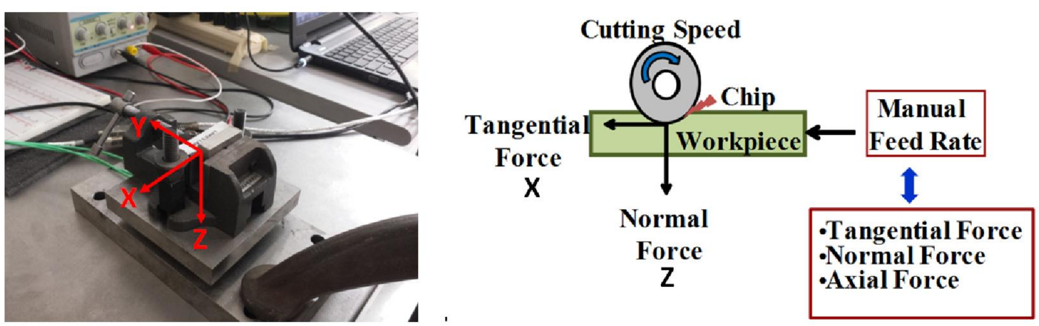

Figure 2: An image of a grinding sample (left) and a figure showing the force data collection module (right). Forces in three directions were measured, tangential (x-axis), normal (z-axis) and axial (y-axis).

The material used in this study was grade 304 stainless steel in form of test coupons with dimensions of 4.5cm in length, by 1.5cm in width, by 2cm in height. The bulk material was received as annealed and roll leveled. At room temperature, grade 304 stainless steel shows austenitic properties.

The grinding experiment was conducted with a Dremel 4000 hand held power tool using alumina sanding bands of 240 grit sizes (mesh number). The power tool was running at a constant speed of 5000 rpm. All grinding operations were performed under dry cutting conditions. The grinding force was varied manually which produced force variations in the tangential, normal, and axial directions as shown in Figure 2. A piezo-electric transducer based load cell (Kistler 9252A) was mounted under the workpiece to measure the grinding forces during machining. A vise was used to clamp the workpiece to the sensor. The force data were sampled at 1000 Hz using a National Instruments DAQ and Labview software.

2.2 Procedure

The experiment proceeded in the following manner:

- Before each grinding trial, a calibration step was carried out in order to collect the particular fixational and saccadic statistics of each subject (rules from classifying a state of gaze as fixation or saccadic will be discussed in Section 3). Each subject was asked to fixate on separate points located on a chart approximately 1450mm in front of the glasses, a distance which corresponds to the position of the gaze plane. After fixating on a single point for five seconds, the subject was asked to shift their gaze to another point and fixate for another 5 seconds. This process was repeated two times.

- Each of the subjects was asked to grind the material for a duration of 1 minute. They were instructed to make the piece as smooth as possible. The gaze data as well as the force data were collected and saved separately for each trial and each subject.

- Both the grinding wheel and the grinding sample were replaced for each trial.

- Each subject performed three trials.

2.3 Remarks

A major challenge in any experiment of this type is the faithful capture of natural behavior without the impediments of an artificially constructed environment or the presence of interfering measurement equipment. Each subject must be permitted to use their sensorimotor skills in their own unique ways pertaining to their level of experience. A complete elimination of these confounds is difficult to achieve, and we made every effort to design an experiment that afforded the subjects the use of their sensorimotor skills as naturally as possible. For the series of trials, each subject had an introduction to the grinding protocol as well as a single practice run before the data was collected.

3 Data Processing Methods

In this section we describe the basic data processing methods that are common to both of the techniques described in the next two sections. We choose our methods from established procedures with minor modifications, bearing in mind that we are interested in examining gaze-motor characteristics of our subjects and how they may differ based on skill.

3.1 Scanpath

A scanpath is the trace of the subjects eye movements in space and time (Duchowski, 2007), (Cristino et al., 2010), (Noton and Stark, 1971), (Cristino et al., 2010), (Holmqvist et al., 2011). It is a locus of points *(x,y,t)* which describe when and where the subject attends to a particular visual stimulus. True eye movement is complex and continuous. The scanpath is the distillation of a point of regard from the general eye movement. The particular choice of presentation of the scanpath varies by the type of analysis. Among these presentations, the most common consists of plotting the x and y coordinates of each fixation point onto an image of the visual stimulus (Holmqvist et al., 2011). The duration of the fixation is illustrated using a circle with a diameter proportional to the amount of time.

In order to generate our scanpaths, we distilled fixations from the time history of the BPOR using a heuristic that encompasses common practices established in (Holmqvist et al., 2011). The strength and weaknesses of these methods have been explored in (Salvucci and Goldberg, 2000). To produce good data, the SMI gaze tracker requires an unoccluded view of the pupil. The dominant periodic occlusion events are blinks and these are manifest in the data as points at the origin. However, long eyelashes, drooping eyelids, or any other occlusion are generally identifiable by spurious breaks in the continuity of the data. These discontinuities must be removed before generating fixations from the data.

It is generally accepted that the eyes cannot move faster than a given speed; usually 900 degrees per second. It is also generally accepted that saccades can be defined as shifts that occur above certain speeds. Furthermore, the human attention system cannot interpret complex visual stimuli lasting for a duration of less than 200ms (Holmqvist et al., 2011). Therefore, we used these physiological limitations as a means of classifying the data as a saccade or a fixation.

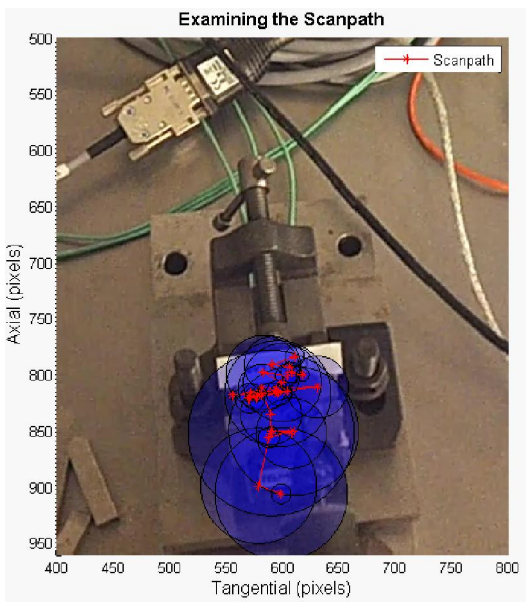

Figure 3: An example scanpath superimposed with an image of the grinding sample. The centers of fixations are denoted by stars. The durations of fixations are represented by the diameters of the circles. The fixation centers are connected by straight lines according to their temporal order. Each straight line corresponds to a saccade.

The steps of the heuristic are as follows:

1. Calculate the angular velocity of the eye and remove points with unusually high velocity (< 900 degrees/sec).

2. Interpolate between the removed points to create a smooth signal.

3. Remove any remaining points located at the origin (as a result of blinks or loss of pupil tracking) and those points located outside the data window.

4. Calculate the angular velocity of the eye. Velocities above a threshold of 10 degrees/sec are saccades, while those below are fixations.

5. Compare the fixation durations against a threshold of 200ms. If the duration is less than the threshold, disregard it.

6. Calculate the mean value of the position of the eye for each fixation. The resulting vector tuple is the *(x,y,t)* elements of the fixation. Transitions between fixations are then saccades.

An example scanpath of one subject and one trial is shown in Figure 3.

It is important to note that gaze data acquired from head mounted eye trackers cannot distinguish between "eye-in-head" fixations in which the eye moves from one stimulus to another, and "eye-on-stimulus" movements in which the head moves with respect to the visual stimulus while the eye remains fixated on it (Duchowski, 2007), (Holmqvist et al., 2011). Elimination of "eye-on-stimulus" events would necessitate the use of a fixed-head gaze tracker with a much higher sampling rate. The use of a fixed-head tracker would encumber the subjects in a way that would prevent the capture of their natural behavior. Only through direct observation of the stimulus video are we able to qualitatively detect such differences.

3.2 Filtering of Force Data

The force data were recorded for the normal, axial, and tangential directions. These data were post-processed with a 14th order Butterworth low pass filter at a corner frequency of 5Hz. Normal and axial force data were particularly noisy necessitating the use of such an aggressive filter to extract characteristics at the lower frequencies. Filtering was accomplished using the Matlab function "filtfilt", a bidirectional, zero lag implementation of digital filters.

4 Gaze-Motor Data Visualization

Many of the dynamic aspects of the visual process can be lost when viewing the scanpath as an agglomeration of fixations and durations. For example, the order of fixations become ambiguated without labels, and their directions are not clear without arrows. If the number of fixations is high, or their spatial dispersion is low, gaining insight using some of the traditional presentation methods can be difficult. Figure 3 illustrates some of these visualization problems using data from one of our grinding trials. The primary stimulus is the small grinding sample located near the bottom of the view. Fixations are distributed over a small region and the larger duration circles occlude the smaller ones. It would appear that much of the time is spent fixating on a few select points; however, without another means of viewing the data, we cannot be certain. Likewise, it is difficult to discern the shorter fixations from the longer ones, nor determine when they might have occurred.

A scanpath need not be displayed in a static manner, and many visualization packages will play the temporal evolution of fixations over a video of the stimulus. However, we are illustrating a method to display eye tracking data recorded from dynamic stimuli (grinding task) within a space-time visualization with the focus on the analysis of forces. We are attempting to display the temporal nature of both the attention and motor modalities involved in the grinding task. We have developed a means to present these dynamic data on a single plot which can be used to gain insight into the behavior of the subject.

4.1 Gaze-Motor Space-Time Cube (GMSTC)

Methods for spatial evaluation of data have been developed in the GIScience community to evaluate geospatial trends. Since gaze data is inherently spatial-temporal, some of these methods have been adapted for scanpath visualization (Li et al., 2010). Significant work has been performed which allows the visualization and comparison of visual data between subjects subjected to dynamic stimuli (Kurzhals and Weiskopf, 2013). In general we can use the third dimension to illustrate temporal changes in x and y data. The means of viewing data in this matter is known as the space-time-cube (STC) (Hagerstrand and others, 1968). A scanpath can be extracted along the time axis in order to examine the temporal and spatial shifts in gaze. This is a spatio-temporal approach that helps to elucidate how the eyes explore the visual space as time progresses.

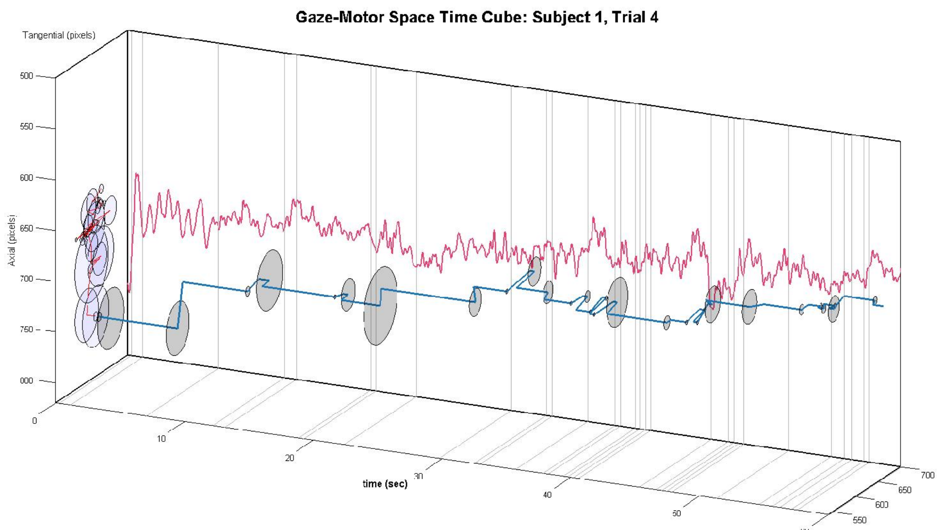

Figure 4: An example of the gaze-motor space-time cube generated from a trial of the experienced subject. Tangential force data is projected along the back plane of the figure while the traditional display of the scanpath is located along the left plane. The temporal evolution of the scanpath is extracted along the time dimension and shown in blue. Gray circles indicate the occurrence of each fixation.

The walls of the space time cube provide ample space to display any other form of temporal data. In our case, we have chosen to display the forces for direct comparison with shifts in gaze. We can create a Gaze-Motor Space-Time Cube (GMSTC) that displays all of these data in a single figure. Figure 4 illustrates an example of this visualization. The scanpath is extruded along the time axis. Changes in gaze position and the order in which they occur are now readily apparent. Each fixation is again noted with a circle whose diameter is proportional to the duration. Each saccadic event is marked on the vertical and horizontal planes. Finally, the tangential force is displayed in the y-t plane.

4.2 Results

Using the GMSTC, we are able to view many of the visual trends that were occluded in Figure 3. In addition, the power of this type of display becomes readily apparent. We are able to determine not only what the subject was looking at and for how long, but also what force they were applying over the duration of a fixation. By examining how the applied forces change with time and the shifts in gaze that accompany them, we can get a qualitative feel for the gaze-motor characteristics of the subject. Figure 4 shows data taken from the experienced subject. The task begins with long fixations and a relatively constant force in the tangential direction. However, after approximately 30 seconds the frequency of fixations increases, and this is followed by an increase in applied forces in the tangential direction.

5 Characteristics of Manual Grinding Skills

In this section, we use the GMSTC to help detect differing characteristics between subjects of different skills. We compare the duration of fixations, the variance of fixation positions, the characteristics of the forces applied and identify a general correlation between the eye movement and force. As noted earlier, we shall refer to the experienced subject as Subject 1 and the intermediate as Subject 2. Furthermore, for the purpose of direction consistency, we will call the *x* direction of the gaze as tangential and the *y* direction as axial. Two videos, one of Subject 1 and one of Subject 2, are available at `https://youtu.be/E5qlCDRXk7E` and `https://youtu.be/z4jP666F_UA`, respectively.

5.1 Fixation Duration

The GMSTCs of two grinding trials, as shown in Figure 4 and Figure 5, show clear distinctions between the fixational duration of the two subjects. A two sample t-test and a chi-squared test were performed on the fixation data. These tests both indicate that trials between subjects are statistically different ($p < 0.03$ for both tests), while the data within each subject are not. This would make sense as we may guess that the fixation durations are an inherent property of the individual subjects. As a result, we pooled the data for each subject across all three trials.

The pooled distributions as illustrated in Figure 6 are highly skewed. A comparison between the two subjects illustrates that Subject 1 in general has longer fixations with a wider distribution of fixation duration. Subject 2 on the other hand has much shorter durations indicating a more frequent shift in gaze. We cannot speak to the importance of any point of regard for the particular subject per se. However, the longer the duration, the fewer the number of fixations that occur in a given trial. Also, a longer fixation implies more time that the subject has to attend to the information from that location in space. So we may imply a greater importance to points of regard that occur at long intervals.

We can invert fixation duration in order to get a distribution of fixation frequency as shown in Figure 7. This is not a true spectral distribution--one could argue that this is a somewhat artificial construction as the fixations defined in analysis are determined as quick shifts in gaze rather than smooth movements. However, we use this method of viewing the data to compare against the spectral distribution of the tangential forces as we shall see shortly.

5.2 Tangential Fixational Variation

Figure 8 shows the distribution of the tangential gaze variations. There are two main reasons for choosing the tangential direction only. First, the task was performed primarily in the tangential direction. Second, there was little variation in the axial direction that could be accounted for as "eye-in-head" movement. As shown in Figure 9, each subject had a tendency to move their view axially as the task progressed thereby mitigating the importance of shifts in that direction--we couldn't tell them apart from re-alignments of the head.

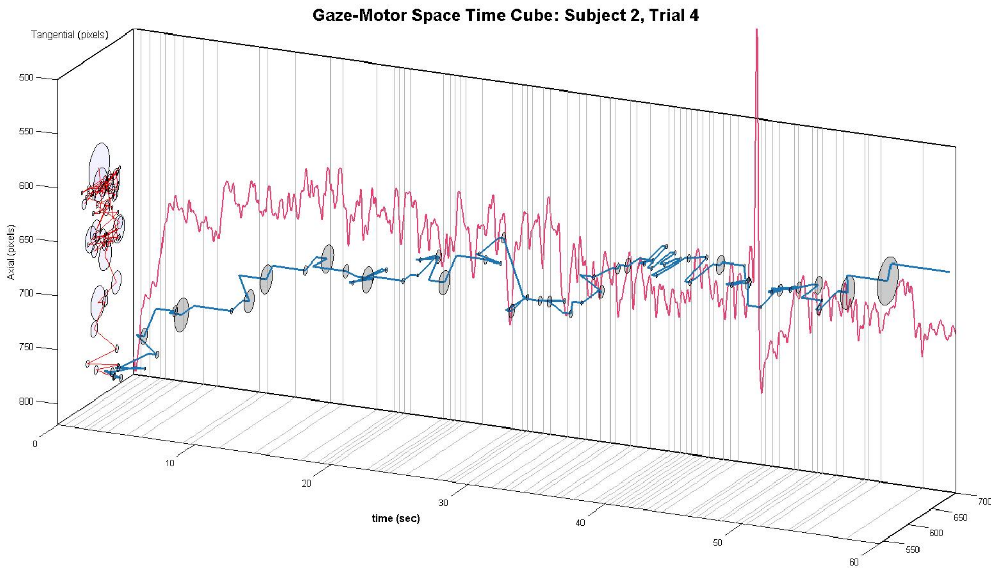

Figure 5: The GMSTC of a trial of the intermediate subject. The figure can be viewed in comparison with Figure 4 which showing the GMSTC of the experienced subject.

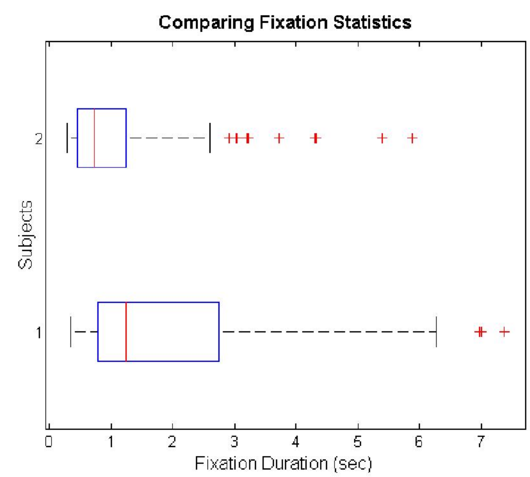

Figure 6: Comparison of skewed fixation distributions between subjects. Data from all trials for each subject is pooled. Red lines indicate the median of the distribution. Whiskers extend out to the 90th percentiles. Outliers beyond the 90th percentile labeled with a red cross.

We calculate the variations as follows. First, the mean for each trial is calculated. The variation is defined as the square root of the squared difference of the gaze position from the mean for each fixation. Thus it is not the variance of a distribution, but the absolute value of the distance of each fixational point of regard from the mean of each particular trial. These data are presented as pixels in the image frame (1280 horizontal by 960 vertical). These data are binned and the distribution of each subject for each trial is plotted in Figure 8. We can see that there is a general trend of lower variation in the more experienced subject (Subject 1) than the intermediate subject (Subject 2).

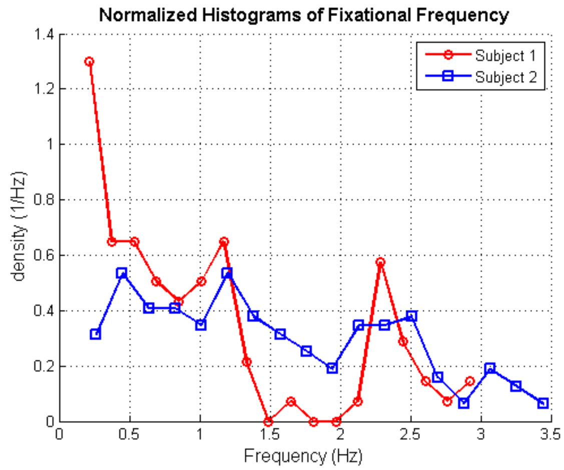

Figure 7: Normalized Histograms of fixational frequency for each subject.

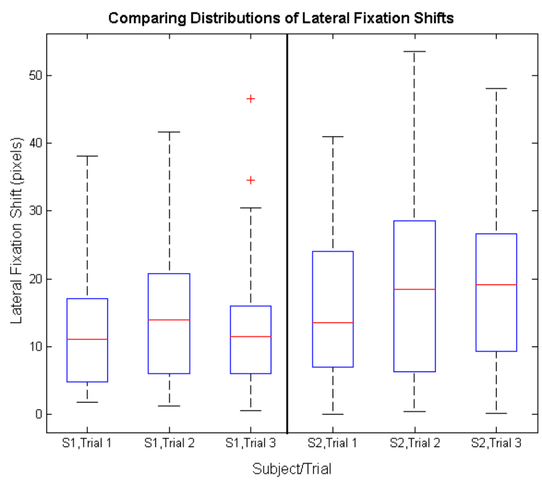

Figure 8: Distributions of the tangential variations for all the trials. The whiskers extend to the 90th percentile of the distribution. Outliers are represented as red crosses.

5.3 Tangential Forces

Figure 10 (a) illustrates the normal versus the tangential forces for a single trial of each of the two subjects. While the normal forces might be considered those that are directly applied by the subject, we can see that there is a strong correlation between the tangential and normal forces ($r^2 = 0.85$). The tangential force is also an indication of the grinding power as the greater the magnitude, the more energy involved in the material removal (Odum et al., 2014). Finally, there is variability in the distributions of the tangential forces between subjects that were large enough to analyze statistically. Therefore, we chose to analyze forces in the tangential rather than the normal and axial directions.

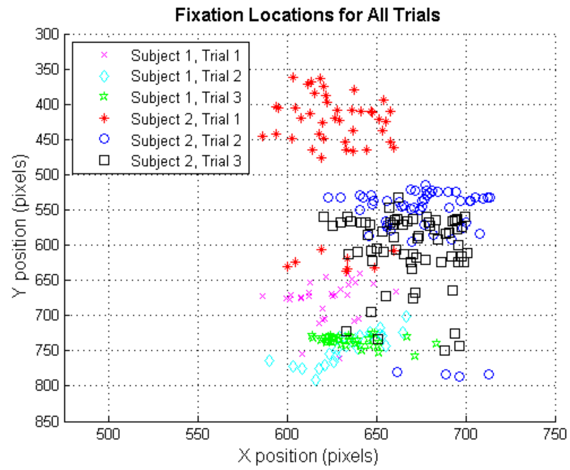

Figure 9: Plot of all fixation points for all trials. Positions are reported in pixels on the original 1280 by 960 pixel field of view. The origin is located in the upper left hand corner of the image frame. Notice the shift in the means of trials can be due, for instance, to the relative position a subject stood with respect to the grinding sample.

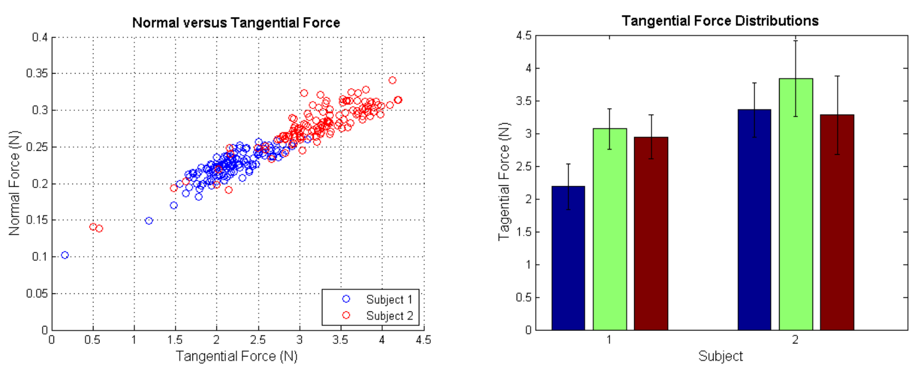

Figure 10: (a) Plot of normal and tangential forces from a typical grinding experiment. (b) Tangential force distributions between subjects. The height of the bars represent the mean force for each trial with the standard deviations indicated.

Figure 11 shows the power spectral density of the tangential forces for the two subjects. These densities are averages across the three trials for each subject. We can see that not only did Subject 2 apply a higher static force, as shown in Figure 10 (b), but these forces contain energy dispersed over a broader bandwidth. This is especially true for frequencies above 1.5Hz.

The modes in these spectra are located at: Mode 1: 0.22Hz; Mode 2: 1.2Hz; Mode 3: 1.6Hz; Mode 4: 2.5Hz. These modes arise through the proprioceptive interaction of the human subjects with the natural dynamics of the mechanical system. We cannot say for certain which regime may predominate in this particular frequency band. However, the clamped test article is extremely stiff, and the grinding wheel was rotating at 5000 rpm. Mechanical resonances are most likely absent at such frequencies. It would not be imprudent to say that the force response in the 0 to 5Hz bandwidth is dominated by the characteristics of the gaze-motor system.

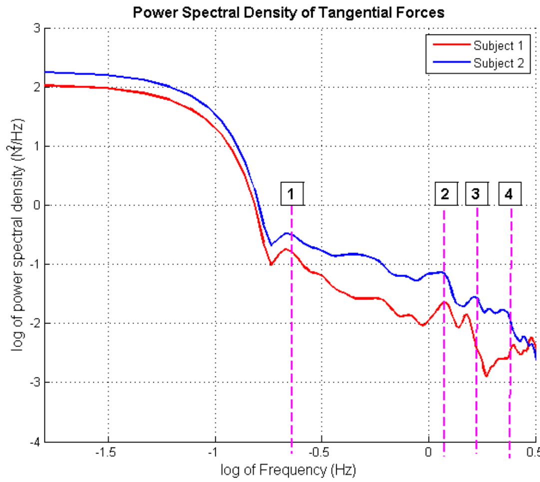

Figure 11: Power spectra of tangential forces averaged over all trials for each subject. Data are plotted between .001 and 5 Hz. The modes of interest are labeled 1 through 4.

5.4 Correlation Between Gaze and Applied Force

In this section, we examine how a subject's shifts in gaze can be correlated with shifts in applied force. Our experimental setup cannot measure where in space a force was applied, only its components along the principle axes of the triaxial load cell. However, we can track how the tangential force changed, and correspondingly, how the eye movements change in the same direction. These corresponding properties are directly evident from an examination of the GMSTC. A close inspection of Figure 5 reveals an interesting quality in the gaze-motor behavior of Subject 2. We can see that the fixations proceed along with, and in relative phase with, changes in the tangential force. The eyes are moving along with the hand as it presses down into the sample.

As a means of comparing the frequency characteristics of the gaze and tangential force properties of each subject, we have overlaid the frequency distributions of the fixations and power spectral density onto a single plot. These plots are shown in Figure 12. We can see from Figure 12 (b) that Subject 2 has a peak in the force spectrum at Mode 2, which overlaps with a similar peak in the fixation distribution. Mode 2 corresponds to the back and forth sweeping motion of the hand and is the primary motion that is visible during the performance of the task. The alignment of these peaks confirms our observation from the GMSTC. There are also, however, overlapping peaks at Mode 3, implying a correspondence between gaze and force at that frequency as well. Similarly, Subject 1 also displays an overlap of peaks at Mode 1. In contrast, there are a number of peaks in the fixation distribution that do not seem to correspond to any modes of the tangential force. These are labeled in Figure 12 (a) as Mode A and Mode B. The difference between these two figures exemplifies the contrast in performance between the two subjects. Subject 1 applied less force, with fewer changes in tangential fixation than Subject 2. Likewise, Subject 1 was less likely to change fixations beyond the initial sweeping hand modes. The other fixational modes may have importance outside of the gaze-motor feedback loop, or may be indicative of another process not captured in these data. More importantly, the information gleened from the GMSTC could be used to impute an importance to particular saccades in the scanpath, and indicate what eye movements are likely to correspond to significant hand movements.

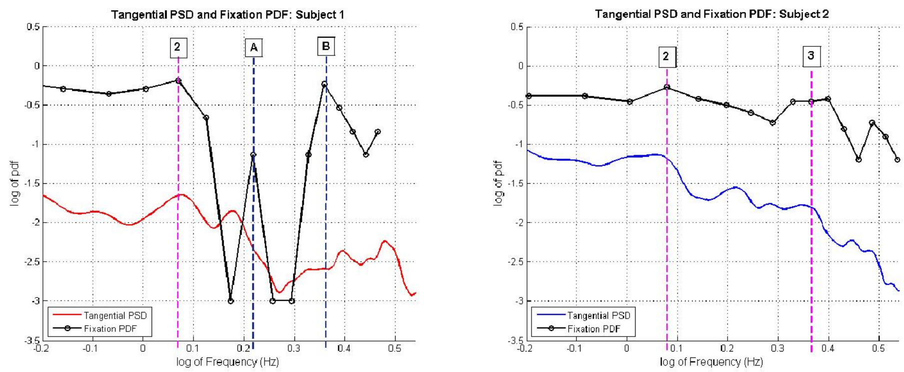

Figure 12: Frequency distributions of the fixations and power spectral densities of tangential forces of (a) the experienced subject and (b) the intermediate subject.

As we have shown, the characteristics of each subject vary. However, both the experienced and intermediate subjects show a correlation of eye movement to force in some degree indicating an importance between the two. This leads for the design of a more thorough experiment that would examine the eye movements for a wider range of tasks (possibly two dimensional grinding) and how the forces change between "important" saccades. The possible rule for quantifying the importance of a saccade has to be a function of the experience of the practitioner.

6 Conclusions and Future Work

Our goal is to develop a principled and data-driven way of analyzing human manual skills. Since the skills involved in complex manual tasks need the close integration of multiple processes, the difficulty for the data analysis then lies on how to deal with data collected from processes of different nature. The paper shows our attempt to tackle such a difficulty. By extending existing methods, such as scanpath methods and space-time cube method, we develop a visualization method with which we are able to illustrate gaze and motor data in a unified representation. Such a representation enables us to carry out meaningful comparisons of data collected from humans with different skill levels. We are currently working on collecting data from a larger and more comprehensive sample population which may also capture confounding effects on behavior such as gender or age. We are also continuing to develop an algorithm to learn a joint gaze-motor model, and study the implications of the learned model to the design of human-robot systems.

References

Bauer, A., Wollherr, D., Buss, M., 2008. Human–robot collaboration: a survey. Int. J. Humanoid Robot. 5, 47–66.

Begum, M., Karray, F., 2011. Visual attention for robotic cognition: a survey. Auton. Ment. Dev. IEEE Trans. On 3, 92–105.

Borji, A., Itti, L., 2013. State-of-the-art in visual attention modeling. Pattern Anal. Mach. Intell. IEEE Trans. On 35, 185–207.

Borji, A., Sihite, D.N., Itti, L., 2014. What/where to look next? Modeling top-down visual attention in complex interactive environments. Syst. Man Cybern. Syst. IEEE Trans. On 44, 523–538.

Cristino, F., Mathôt, S., Theeuwes, J., Gilchrist, I.D., 2010. ScanMatch: A novel method for comparing fixation sequences. Behav. Res. Methods 42, 692–700.
Duchowski, A., 2007. Eye tracking methodology: Theory and practice. Springer Science & Business Media.
Duchowski, A.T., Driver, J., Jolaoso, S., Tan, W., Ramey, B.N., Robbins, A., 2010. Scanpath comparison revisited, in: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. ACM, pp. 219–226.
Erez, T., Tramper, J.J., Smart, W.D., Gielen, S.C., 2011. A POMDP Model of Eye-Hand Coordination., in: AAAI.
Goldberg, J.H., Helfman, J.I., 2010. Scanpath clustering and aggregation, in: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. ACM, pp. 227–234.
Goldstein, E., 2014. Cognitive psychology: Connecting mind, research and everyday experience. Cengage Learning.
Guo, C., Zhang, L., 2010. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. Image Process. IEEE Trans. On 19, 185–198.
Hagerstrand, T., others, 1968. Innovation diffusion as a spatial process. Innov. Diffus. Spat. Process.
Hayhoe, M., Ballard, D., 2005. Eye movements in natural behavior. Trends Cogn. Sci. 9, 188–194.
Hejmady, P., Narayanan, N.H., 2012. Visual attention patterns during program debugging with an IDE, in: Proceedings of the Symposium on Eye Tracking Research and Applications. ACM, pp. 197–200.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye tracking: A comprehensive guide to methods and measures. Oxford University Press.
Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 1254–1259.
Jarodzka, H., Holmqvist, K., Nyström, M., 2010. A vector-based, multidimensional scanpath similarity measure, in: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. ACM, pp. 211–218.
Kang, Z., Landry, S.J., 2015. An eye movement analysis algorithm for a multielement target tracking task: Maximum transition-based agglomerative hierarchical clustering. Hum.-Mach. Syst. IEEE Trans. On 45, 13–24.
Klocke, F., Dambon, O., Behrens, B., 2011. Analysis of defect mechanisms in polishing of tool steels. Prod. Eng. 5, 475–483.
Kong, Z., Mettler, B., 2013. Modeling human guidance behavior based on patterns in agent–environment interactions. Hum.-Mach. Syst. IEEE Trans. On 43, 371–384.
Kurzhals, K., Weiskopf, D., 2013. Space-time visual analytics of eye-tracking data for dynamic stimuli. Vis. Comput. Graph. IEEE Trans. On 19, 2129–2138.
Land, M.F., 2009. Vision, eye movements, and natural behavior. Vis. Neurosci. 26, 51–62.
Lee, J., Bagheri, B., Kao, H.-A., 2015. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manuf. Lett. 3, 18–23.
Li, X., Çöltekin, A., Kraak, M.-J., 2010. Visual Exploration of Eye Movement Data Using the Space-Time-Cube, in: Fabrikant, S.I., Reichenbacher, T., van Kreveld, M., Schlieder, C. (Eds.), Geographic Information Science. Springer Berlin Heidelberg, pp. 295–309.
Mettler, B., Kong, Z., Li, B., Andersh, J., 2014. Systems view on spatial planning and perception based on invariants in agent-environment dynamics. Front. Neurosci. 8.
Noton, D., Stark, L., 1971. Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vision Res. 11, 929–IN8.
Odum, K., Castillo, M.C., Das, J., Linke, B., 2014. Sustainability Analysis of Grinding with Power Tools. Procedia CIRP 14, 570–574.
Posner, M.I., 2011. Cognitive neuroscience of attention. Guilford Press.

Privitera, C.M., Stark, L.W., 2000. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. Pattern Anal. Mach. Intell. IEEE Trans. On 22, 970–982.

Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. Q. J. Exp. Psychol. 62, 1457–1506.

Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols, in: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications. ACM, pp. 71–78.

Siagian, C., Itti, L., 2009. Biologically inspired mobile robot vision localization. Robot. IEEE Trans. On 25, 861–873.

Sprague, N., Ballard, D., Robinson, A., 2007. Modeling embodied visual behaviors. ACM Trans. Appl. Percept. TAP 4, 11.

Todorov, E., 2004. Optimality principles in sensorimotor control. Nat. Neurosci. 7, 907–915.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., Crundall, D., 2003. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics 46, 629–646.

Wojtara, T., Uchihara, M., Murayama, H., Shimoda, S., Sakai, S., Fujimoto, H., Kimura, H., 2009. Human–robot collaboration in precise positioning of a three-dimensional object. Automatica 45, 333–342.

Yi, W., Ballard, D., 2009. Recognizing behavior in hand-eye coordination patterns. Int. J. Humanoid Robot. 6, 337–359.