Swansea University

Computer Science

Applying K-Means Clustering to Photon Patterns for an Optimal Approach to Photon Density Estimation

Abstract

A presentation of our investigation to discover a new method in delineating photon density estimation in the form of a consistent procedure that operates efficiently in place of our linear and conventional approaches. We analyse current viable methodologies including scenarios that involve improved progressive stochastic photon mapping and we investigate our chosen method of querying our photon map with an implicit k means algorithm.

Author:
Fadl Akkad
799347

Supervisor:
Dr. Benjamin Mora

May 2017

Department of Computer Science
Swansea University

Contents

1 Introduction

Global illumination is a subject in computer graphics that has consistently proved to yield fruitful results with further research and innovation, yet there is still room for refinement. Constructing a realistic and immersive digital environment is the imperative goal in this field of study and ensuring our scene has naturalistic and authentic lighting is a distinctly fundamental element needed to be taken into consideration throughout the rendering process. In the topic of illumination we are concerned with having lighting behave in a real to life manner in relationship to the originating light source. Where there is an object obstructing illumination from a light source, we desire a shadow in the resulting area, we plan to have various algorithms dedicated to simulating indirect light paths including specular or caustic reflections where applicable, where there is colour, we expect there to be an absorption or inter-diffuse reflection according to real world physics.

Today there is a multitude of methods used to bring about immersive illumination. A common process is through ray tracing, a well documented algorithm that also has many applications outside of the study of illumination. This process involves projecting a 'ray' into the scene in order to trace and detect legitimate paths from a contacted object to a light source. This ray has its own flexible actions and responses in the case of making contact or 'collision' with objects including reflections and refractions in order to reach a light source. The implication from this is the further elaborated recursive ray tracing; where one ray may have opportunities for multiple collisions with various objects in the environment making for an iterative procedure. However in recent times, ray tracing has been been decided to not be the most optimal method of illumination, researchers and developers instead prefer utilising approaches concerned with 'global' illumination rather than 'direct' illumination, which neglects taking into consideration indirect light paths.

Global illumination on the other hand accomplishes emulating the effects of indirect light paths adequately although at the expense of computational power. There are various workable methods utilised in order to accomplish global illumination. Path tracing; an unbiased procedure that applies Monte Carlo Integration once contact is made within the scene proves to be advantageous, we use Monte Carlo Integration to sample a number of directions in a hemisphere around the point. This will help depict certain details such as colour, the more samples the more accurate the image will be. However such methods come with their own disadvantages, while path tracing continues to be advantageous for developers it proves to be very expensive in terms of computational time.

On the other hand the process of mapping photons in the scene and subsequently using that map to delineate density estimations proves to be less expensive and a common method amongst developers. To elaborate, a large number of photons are dispersed from a light source, these photons behave flexibly like a ray would, where these photons eventually 'land' or settle in a point is registered in a super imposed map, from here we may send a ray to make contact and deploy a variety of algorithms such as applying a kernel or applying 'K-Nearest Neighbour' to measure how densely populated the immediate proximity is, this then helps us measure how intensely luminous an area should be although not necessarily accurate due to included bias. In recent years a great deal of research has been made in this subject, providing for such methodologies as progressive photon mapping capable of computing lighting effect without storing photons, while advantageous this approach also has its own constraints.

We acknowledge that methods associated with global illumination can be generally cost effective during the rendering procedure, consequently we are interested in the area of research that seeks to discover a more improved and optimal way of deploying realistic indirect lighting, ideally in such a fashion as to lose dependency on any such linearly selected method and instead explore an efficient and universal procedure that can be applied to any situation uniformly to provide a systematic and iterative rendering experience to be applied to a given scene. Due to the nature of our investigation, we subsequently explore various different possible approaches such as progressive stochastic photon mapping (sppm) and the technical field of machine learning and what it has to offer us as well as why it may be the best viable approach. This paper serves to further explain our investigation including our methodology, development and application.

2 Motivation for K-Means Clustering

We are concerned with arising problems encountered in the biased nature regarding the photon mapping - density estimation procedure, thus our task is to discover a distinct methodology that will allow the development of an iterative process for handling efficient density estimation that can be applied to any given scene.

A typical attribute of any field of research regarding computer science is the trade off between speed and accuracy. It is imperative to examine the consequences of the algorithms we deploy, including the possibility for counter productive computational time and also severe exhaustion of computational power, a clear example of this is applied path tracing. Also when utilising any given method for global illumination we find that we are constrained by certain limitations, for example the dispersal and subsequent attempts at estimating the density of populated photons can not be considered reliable or effective in all cases, this is because there is a possibility that too much of a blurring effect is detected in the sample area due to a high number of k sample points, while there is also a chance for noise or variance that can occur using too low a number of k. At times due to complex scenarios of varying diffusions, reflections and refractions, the effect of the light emitted by the photon is not as intended, thus we acknowledge that it is best to be careful in determining the reliability of the lighting we visualise.

In a sense, since it is our understanding that our task is to better this method we naturally begin to explore the possibility that our objective is to critically examine and look for an alternative way of constructing global illumination using dispersed photons. There has been a multitude of research regarding the improvement of the photon mapping - density estimation procedure that requires examination including progressive stochastic methodologies calculating correct luminosity radiance, however it is common to discover photons being unevenly distributed due to occlusion causing certain regions to have relatively low photon density, leading to detrimental experimental results. This paper shall examine such methods, their obstacles and research related to them.

We affirm that it is nominal to alter conventional methods revolving photon density estimation and investigate an elaborate algorithm that is able to better analyse the behaviour and characteristics of photons in real time in order to delineate expected consequential effects. Our line of investigation therefore leads us to the ability of better enhancing illumination effects in the scene, hence we consider exploring the technicalities of the machine learning K-Means sort and what such a methodology has to offer us.

Machine Learning algorithms are formulated to provide an elaborate procedure which allows our machines to be able to identify unique patterns through its experiences with previously learned and predetermined data samples that exhibit similar patterns. The K-Means sorting procedure proves to be advantageous in the scenario of density estimation in general, and thus naturally we inquire as to what it offers graphical environments when a photon map is queried.

We observe then that our task is to apply the K-Means technique to our rendered scene in order to deliver density estimation for evaluation. Our theory is that once we generate a large number of photons and subsequently have them disperse from distributed light sources, we collect characteristics of the photon settlement points through querying the photon map and experimenting with our algorithm.

To better formulate our research question; We ask if there is a way to create an enhanced procedure towards photon density estimation that operates efficiently and effectively in place of conventional approaches by exploring a clustering sort designed to delineate implicit effects and consequences from an extensive set of samples.

3 Related Work

We analyse and identify the advantages and obstacles featured in the methodologies currently used for global illumination and density estimation and provide insight on how a more enhanced clustering technique could prove to be more effective.

3.1 Examination of Viable Global Illumination Methodologies

There has been a multitude of studies and enquiries in the subject of global illumination and how we can further refine and discover the methodologies we use in order to realise our desired effects. A clear example of such study is the innovations regarding path tracing and Monte Carlo rendering. A paper from Doidge et al. [DJM12] expresses an approach designed to eliminate the noise caused by caustic paths during this procedure. The algorithm developed achieves far less noise from diffusely lit regions, whilst producing optimal caustic effects.

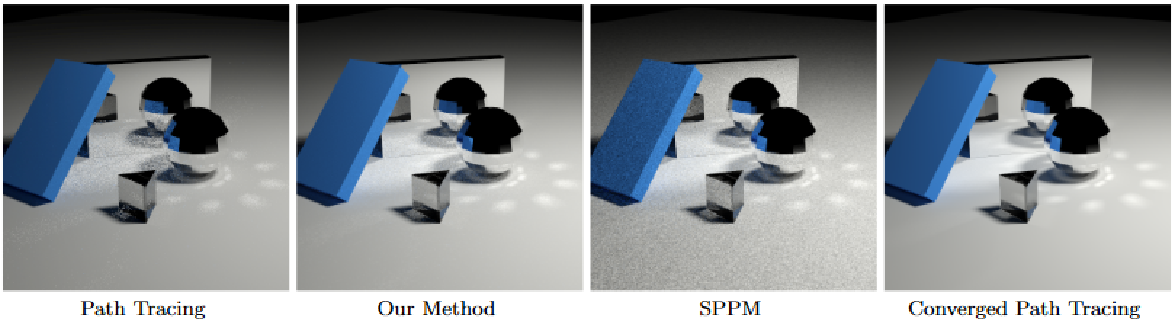

Figure 1: *Progress made by Doidge et al. "Mixing Monte Carlo and Progressive Rendering for Improved Global Illumination"*

In addition to this Veach and Guibas [VG95] discuss alternative approaches to Monte Carlo estimators by combining samples from several distributions. Both these studies suggest that reducing the time taken for path tracing to achieve its task proves to be of a paramount importance for global illumination efforts. It is important to note that Lafortune and Willems [LW93] in their paper discuss at length the application of bi-directional path tracing, whereby contact points on the particle paths are connected using 'shadow' rays and subsequent further details are provided. Bi-directional path tracing constitutes a good example of the use of photon trajectories as an effective tool to provide better optimisation and hence path tracing is important to consider in our investigation due to its use of photons as a critical factor in constructing enhanced global illumination. We observe the benefits of having a more elaborate randomised operation of photon dispersal and radiance calculation, however we also observe that drawbacks such as noise and oversampling remain unavoidable as well as the large amount of storage needed for such lighting effects. Since we are testing for speed and efficiency, it may be cost effective to lose dependency on Monte Carlo rendering as a sole solution.

Various applications have gained success from using prediction and estimation algorithms such as visual and audio identification, and the reason for this success is attributed to the extensive data sets consistently used by the algorithms which in turn aggregates more optimal results. Hence our theory is that the more K-Means clustering is used, the more extensive the algorithm will be in determining photon density involved ina scene, which is a major advantage to using K-Means for the effort of instantiating a more beneficial procedure that can be applied to changing scenarios. In this light K-Means clustering may provide the potential for identifying photons by category for relaxation purposes at an efficient rate while maintaining the quality seen from the use of SPPM or better.

Further more, there are multiple unique studies dedicated to global illumination, one such example is the use of multiple sets of virtual directional light or octrees as explored in a paper by Vivanloc et al. [Viv+07] The methodology discussed is a great contribution towards an effective rendering approach in real time; assorting photons into K-Means clusters so that they may be utilised for allocated divisional calculations is a common approach amongst developers and this method shows us that developing to minimise excess storage and time is critical. Another example; Budge et al. [BAJ08] discusses a novel approach that generates photons that interact specifically with specular geometry, these photons are then divided throughout the scene, allowing for equally effective caustic effects in the various regions of the environment although demonstrates that details such as caustic lighting cannot be computed if the originating light source is a point light and calls for the need for unbiased effects to be more successful in their outcomes, this may be solved by exploring learned prediction. Dammertz et al. [DKL10] provides an algorithm that recursively uses a memory footprint for performance without the need for pre-computations giving a efficient rendering solution independent of specific parameters.

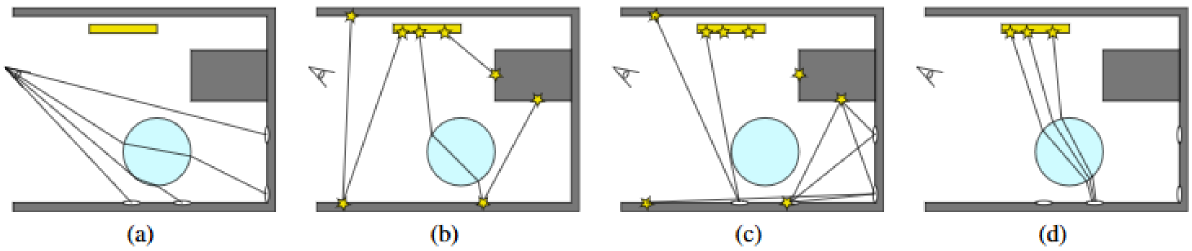

Figure 2: *Diagram presenting Dammretz et al's. portrayal of eye path vertices. "Progressive Point-Light-Based Global Illumination"*

We thus see the potential advantages for a more efficient clustering algorithm that allows for recursive query's to be performed in real time.

We affirm that in our investigation it is important to consider processing power and how to better maximise efficiency and bypass time constraints. Arthur and Vassilvitskii [AV06] discuss the speed of the K-Means application, implying the worst case running time is superpolynomial, this and other algorithms are further explored by Wang et al. [Wan+09], Guatron et al [Gau+05] and Benthin et al [BWS03] who discuss at length the complexities and objectives related to maximising GPU efficiency during the global illumination process. These studies demonstrate the that during our experimentation with K-Means clustering we will we need to retain a cautious approach as to how well our program performs in terms of speedup and how we can minimise potential GPU computation excess.

3.2 Innovations Regarding Stochastic Progressive Photon Mapping

As previously indicated there has been numerous contributions to the improvement of the photon mapping - density estimation procedure as fully expressed by Jensen [Jen96]. Additionally Walter [Wal98], Shirley et al. [Shi+95] and Myszkowski [Mys97] offer contributions to how we understand and approach density estimation. We acknowledge that one of the most nominal strategies for global illumination is progressive photon mapping. Hachisuka et al. [HOJ08] present and discuss ppm and its role in the evolution of global illumination. The paper demonstrates the evolution of methodologies concerned with global illumination including the stark contrast between path tracing and more concurrent strategies which prove far more beneficial.

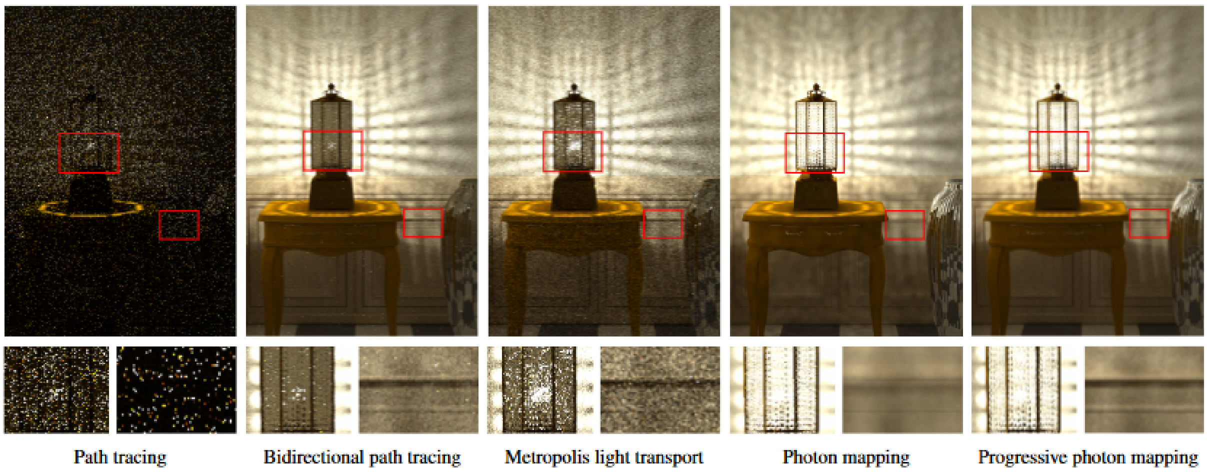

Figure 3: *Diagram by Hachisuka et al. demonstrating the results of PPM "Progressive Photon Mapping"*

We observe that the photon mapping procedure delivers more optimised and authentic illumination effects when applying a multi-pass algorithm where the first pass is ray tracing followed by any number of photon tracing passes. The paper discusses the effects of elaborating on the Monte-Carlo method by using a radiance estimate that iteratively converges to the correct value.

Furthermore Hachisuka and Jensen elaborate in their paper [HJ09] the stochastic innovation. The paper discusses at length the inherent problem related with ppm which is the inconsistency of rendering precise distributed ray tracing effects such as depth-of-field due to the inability to compute the correct average radiance value over specific regions. SPPM proves to be more advantageous as it utilises shared photon statistics within the region rather than isolated photons. Chen et al in their paper [CWY11] further innovate upon sppm when used in conjunction with the metropolis Hastings algorithm which provides effective probability distribution to further optimise the distribution of photons.

These studies show the profound influence that random sampling has on the photon mapping procedure, we also observe in results of experiments that time constraints still proves to be a problem, in fact execution time with these added methodologies has increased to that similar of conventional path tracing. Another problem is present anisotropic errors such as objects at a greater depth of view consisting of inaccurate luminosity effects.

3.3 Closer Inspection Regarding Photon Mapping and Density Estimation

Additionally it is important to consider the research regarding photon relaxation; the critical objective of removing noise from luminous regions. Spencer and Jones in their papers [SJ13b], [SJ09b] and [SJ13a] explore their strategy of reformulating the radiance estimator to accommodate a radiance reconstruction in which error is significantly reduced. This is achieved by using Voronoi diagrams to accurately calculate and adjust settled photons to serve better authentic results, in effect relaxing photons to a more optimal format.

Figure 4: *Progress made by Spencer and Jones demonstrating photon relaxation constraints. "Into the Blue: Better Caustics through Photon Relaxation"*

The results of this are positive as demonstrated on more elaborate, complex illumination effects such as caustic lighting where reduced noise and consistently relaxed photons provides for a smooth and explicit image. Other experiments related to photon mapping prove to be a valuable strategy used to achieve global illumination.

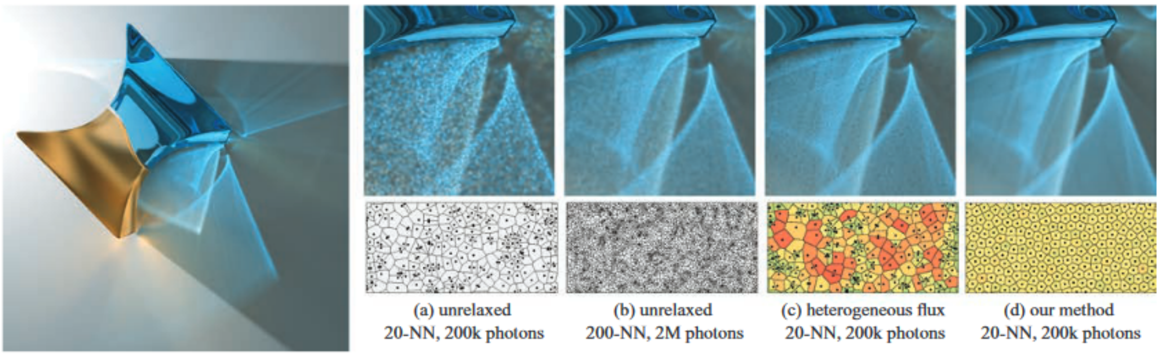

Figure 5: *Progress made by Spencer and Jones. "Progressive Photon Relaxation"*

One such example is from Shregle [Sch03] who provides a bias compensation which seeks to adapt the density estimation bandwidth to a more preferable level. Spencer and Jones [SJ09a] further explore this subject with hierarchical photon mapping, they discuss at length the advantage of applying hierarchical evaluations of photons on each surface as an alternative means of calculating the radiance estimator. In conjunction with this, Keller et al. [Kel+] offers a definitive algorithm to compute simultaneous photon map queries. All of these studies assist us in our understanding of the technicalities involved in the photon mapping procedure and provide effective results that we may compare with our investigation using an enhanced clustering technique.

Additionally Spencer and Jones offer progressive ideas as to how to proceed with querying our photon map, in their paper [SJ09a] they discuss at length their method of hierarchical photon mapping, on the effect of breaking down and segmenting the scene into multiple different and unique zones or regions of which to collect successive photon data from. This is done using the radiance estimator and significant multipliers in order to gain better perspective.

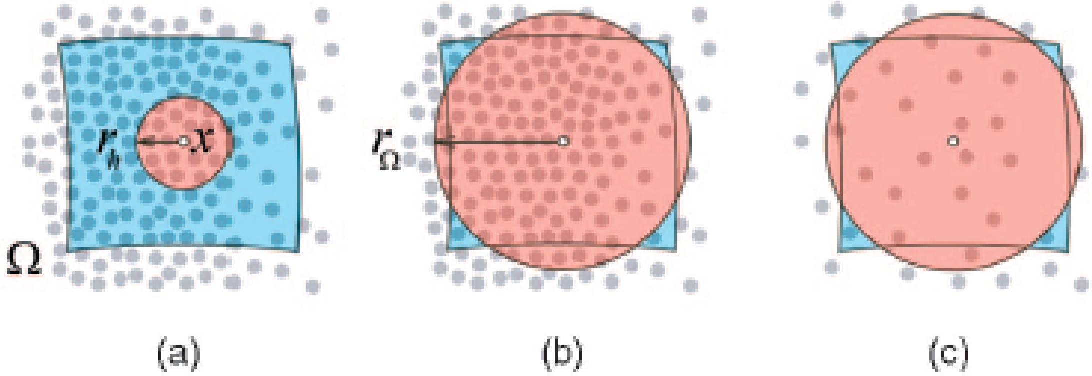

Figure 6: *Demonstration by Spencer and Jones of the more efficient uses of radiance estimates, with (c) being the result of purposeful photon categorisation. "Hierarchical Photon Mapping"*

The paper further investigates enhancing their photon query algorithm by amending their radiance estimation with interpolation calculations.

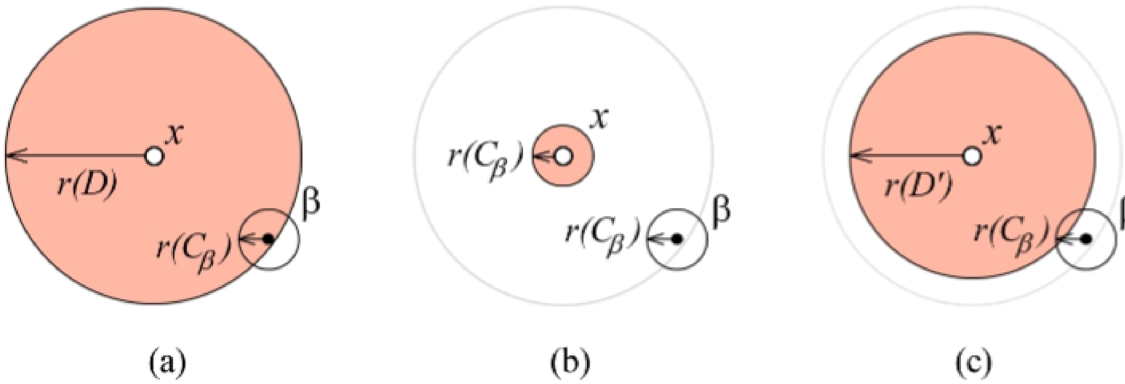

Figure 7: *Spencer and Jones demonstrating their method of coherence interpolation. "Hierarchical Photon Mapping"*

In summary the paper serves to demonstrate a clear methodology among many as to the process of querying our photon map, in our investigation we will be using sequential gathers of our scene while regulating a radiance estimator that will attempt to provide accuracy for implementing scene scenarios in a K-Means application.

4 Research and Methodology

The following presents our understanding of concepts revolving around our investigation and our chosen methodology

4.1 The Efficacy of K-Means Clustering for Density Estimation

Since we seek to overcome the problems associated with current density estimation procedures such as time constraints, anisotropic errors and dependence on random sampling, we apply an unsupervised learning algorithm to discover the results concerned with determining effects based on given patterns.

K-Means Clustering proves to be a rewarding field of study for a plethora of applications regarding computer science both for many academic purposes, particularly involving determining data specifics from samples, the procedure proves useful for constructing a program dedicated to identifying patterns and making decisions on data based on those patterns.

One great benefit of using a unsupervised learning algorithm is that a great number of scenarios can be applied to it and in effect provides a workable degree of flexibility. K-Means clustering proves to be more accurate than a more conventional algorithm such as the KN algorithm, whereby N centroids are populated in the vicinity, from which categorization is determined by majority vote. K-Means clustering takes more authentic variables into consideration.

Such an algorithm includes has historically proven to be dedicated to the assistance of identifying patterns in densely populated samples and we investigate if this algorithm applied to a photon map can accomplish the density estimation phase of the rendering process by determining the characteristics of the lighting based on those identifications. Typically as an unsupervised learning algorithm our investigation will include a collaboration between a rendering program such as POV-Ray and machine learning program MATLAB. We hope to achieve a level of sophistication in the results obtained by K-Means clustering on photon query gathers to evaluate the capacity and effectiveness of the subsequent results.

Our procedure can be summed up by the following:

- Query dispersed photon map through progressive sequence of gathers and respective radiance estimates.
- Identify unique patterns and categorisations of the positions of photons where they settle.
- Record and observe the proximity of photons to calculate the varying degrees of regional light intensity.
- Further apply the algorithm with each additional environment, to examine and compare the effects of different scenarios.
- Evaluate results delivered by K-Means application.

Our goal is to measure the success of this application in a global illumination environment, to discover if a unsupervised machine learning algorithm can help improve efforts towards global illumination effects, particularly density estimation and photon relaxation.

4.2 Producing Sufficient Graphical Environments and Data

Our environments are constructed using POV-Ray which allows for us to generate a large variety of differing and diverse scenes. It is imperative to have unique and varying environments for testing the effectiveness of our algorithm.

The varying scenes will allow for the construction and generation of environments dedicated to providing test cases to sample our algorithm's performance. These environments would ideally include multiple populated objects to varying capacity as well as varying light sources shedding light from different directions for instance. Objects will include speculative and reflective properties, as well as purposeful positioning to project shadows. This will allow us to investigate luminosity effects at greater depth due to the consideration of differing scenarios.

Our algorithm will be formulated using MATLAB, in order to assist our investigation it is important to develop 'artificial' data sets to use along with rendered results that constitute initial testing data, not only will experience gained from scenarios be aggregated but artificial scenarios that emulate differing photon settlement properties will be provided in order to better assist estimation efforts.

When it comes to evaluating our investigation, it is important to provide test cases of the various scenarios so we can demonstrate the effectiveness of the K-Means clustering efforts, hence this paper will document select environments and discuss results attributed to that environment.

During our investigation we retain a rudimentary understanding of a typical machine learning application/experiment.

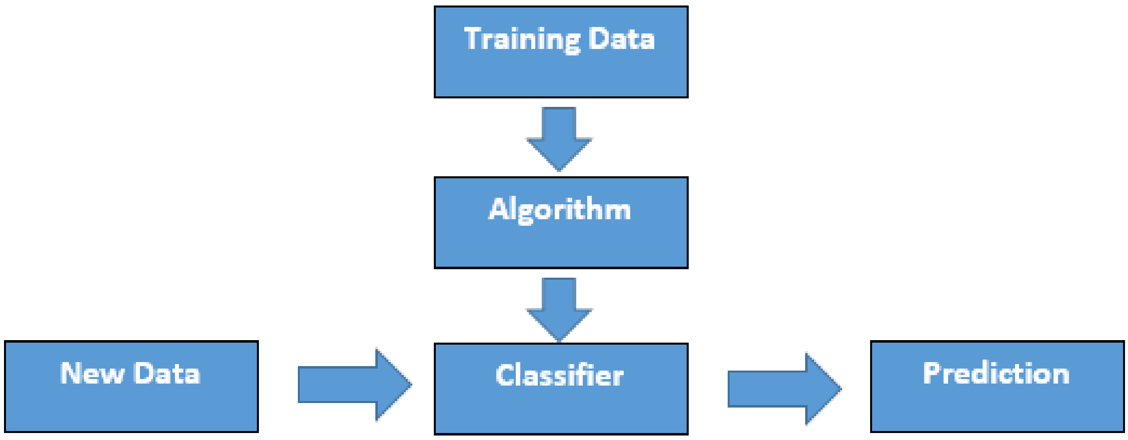

Figure 8: *Diagram demonstrating the standard utility of machine learning*

4.3 Collecting Photon Locations and Categories

A major feature of our investigation is the ability to collect photon locations and subsequently derive necessary categorisations for those photons using our clustering algorithm.

In order for interaction locations to be acquired, our program will be expected to note locations in real time and then work to serialise such data which will act as our feature set for our algorithm. Once our photons have dispersed, our rendering program tracks all photon settlement locations. From here subsequent passes are made through the environment to collect data, this is done using gather and radiance settings, the data is then run through our algorithm to dictate predicted effects and regional intensity. Aside from settlement locations, other characteristics may be taken into account such as origin light sources using colour coding, this will help group photons together by shared features which may prove useful in ensuring fully detailed scene environments. Its important to note the possible excess time and power this may take, especially with potentially hundreds or more photons interacting in the scene all at once, especially in conjunction with the task of predicting for each photon.

Colour categorisation associated with originating light sources proves to be very useful in helping determine where settled photons originate from, further more when accommodating for our clusters, each colour will have a representing number that is assigned to its respective photons.

We affirm that there is a potential risk factor to consider during our investigation. One such example is the possibility that photons may overlap or bundle into dramatically dense populations, potentially causing complications in the overall final result. In order to combat this problem our procedure may have a standardised gather population, this will ensure that our scene is as authentic as possible without any outlier complications.

Aside from this, when it comes to our machine learning application, there is the possibility of errors during the classification process. For example if there is a strong recurrence of a certain location the classifier may complicate the expected result based on an unexpected combination that was not previously learned or exposed. A solution such as extensive training experience with a variety of differing combinations of feature sets is needed in order to overcome this.

From this point on-wards we may test to see whether or not the results are as accurate as intended with close attention to the accuracy of light intensity in densely populated regions.

4.4 Algorithm Fundamentals

The K-Means clustering technique is explained in great detail by Laurence Morissette and Sylvain Chartier in their paper [MC13] where they discuss the main techniques used to implement K-Means clustering.

One such technique is from Forgy and Lloyd, in their respective papers [For65] and [Llo82]. They demonstrate the effectiveness of K-Means when being applied to large data sets, whether that data be perceived as discrete or continuous. The provide the initial basics of the K-Means methodology, including choosing an intuitive number of clusters and the centroids therein as well as the metrics involved in gathering the necessary population for testing. This is then followed by simply applying an element to the nearest cluster.

Additionally Hartigan and Wong [HW79] demonstrate their method which involves a more refined approach to the problem. The algorithm focuses on partitioning of data using sum of squares of errors, which allowed for more flexible selection of data samples that extend beyond standard centroid clusters, with further successive iterations in real time.

Finally MacQueen in his paper [Mac67] provides yet another iterative take on the clustering methodology. The centroids are recalculated every time a case change subspace as well as after each pass through all cases. Centroids are consistently recalculated based on adjusted positioning of the sub spaces. This makes for a more efficient algorithm as it updates centroids more often.

Voronoi diagrams are stipulated by [Vor07] in his paper expressing its technicalities. Perpendicular to the distance between two centroids is a hyper plane that passes through the middle point of the connecting line and divides the space into two separate sub-spaces, thus the general workings of the K-means clustering algorithm.

4.5 Evaluating Results

When it comes to investigations regarding graphical fidelity and particularly in the case of global illumination it is important to consider the details of how test cases are to be examined and subsequently evaluated regarding the comparison of other, more conventional photon relaxation/global illumination techniques.

Our main goal is to test and evaluate the efficiency of K-Means when applied to the scenario of global illumination. We have discussed previously related work regarding efforts to better optimise global illumination, particularly through more refined photon density estimation. Our goal in this investigation is to discover what K-Means clustering can bring to this inquiry. We seek to construct an algorithm that will give results demonstrating a clear comparison with Spencer et al's work regarding photon relaxation for instance, and thus applying the speculated results to ones scene. Due to machine learning contributing to a model that persistently aggregates the algorithm with alternations, this study seeks to discover the capacity to which this can be applied to photon density estimation.

Parameters revolving around our testing efforts will be based on the capacity of regional light intensity; measuring how well a group of photons is illuminated in keeping with how densely a region is populated compared to regions of less photon density. Additionally a more refined methodology will examine the authenticity of the lighting itself in relation to any testing environments and any effect that our clustering techniques has upon that quality and query to what extent our techniques provide to any global illumination alterations within the scene whether subtle or grand.

Our investigation shall conclude with insight gained from our investigation and a judgement of any indication of any success garnered from applying K-Means clustering.

5 Constructing Environments for Photon Map Querying

5.1 Sample Environments

We present the following sample scenes involving photon dispersal that act as test environments for the purpose of our investigation, demonstrating variety in spacial characteristics and object properties. During our investigation scene particulars and object properties are subject to change in order to alter the experiment at hand.

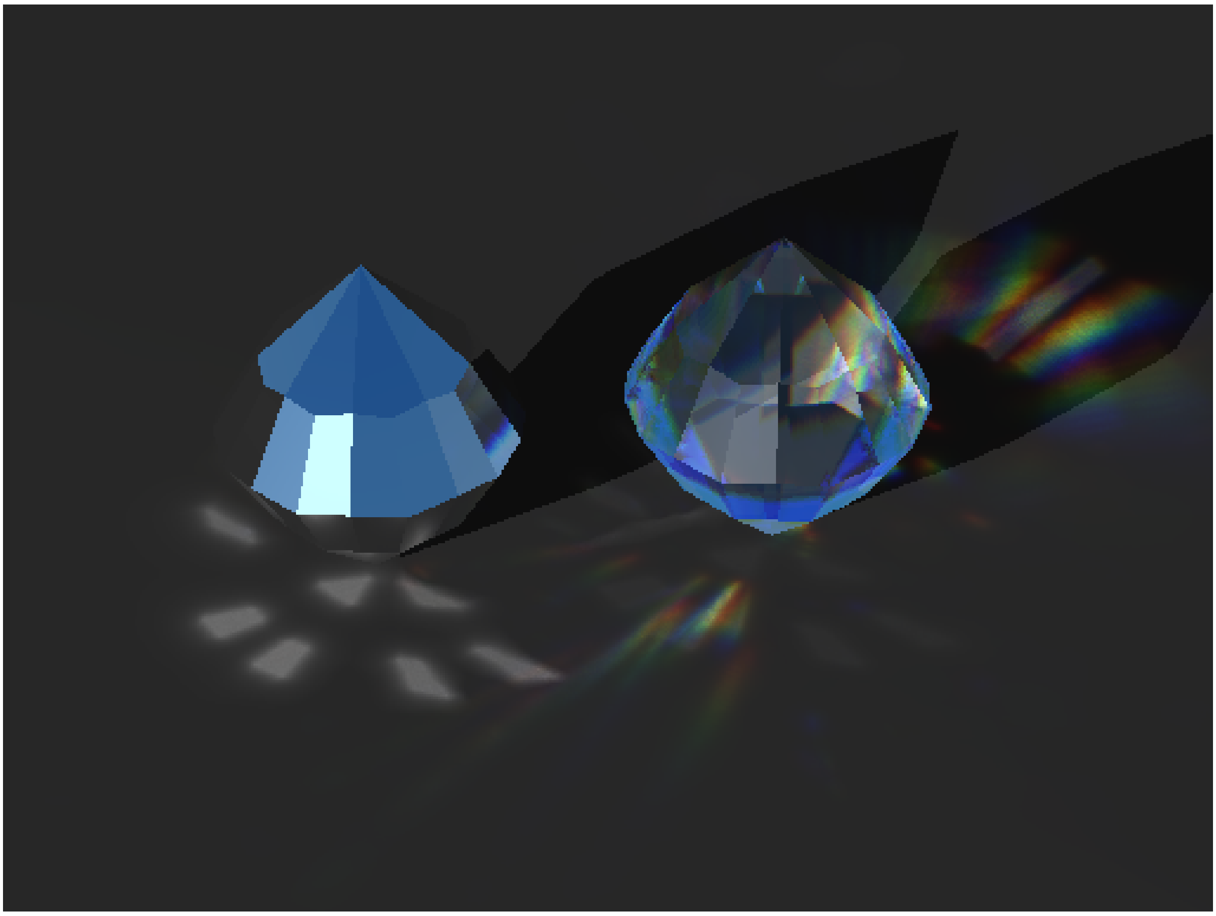

Figure 9: *Sample 1*

A simple environment demonstrating different object properties and the respective effects they have on dispersed photons.

This scene demonstrates a very basic and rudimentary photon dispersal situation. The light source imposes both a shadow on the scene as well as photon initiation. One object is purely reflective, whereas the other object whose respective photons have been designated with colour is a object that has refractive as well as slight speculative properties.

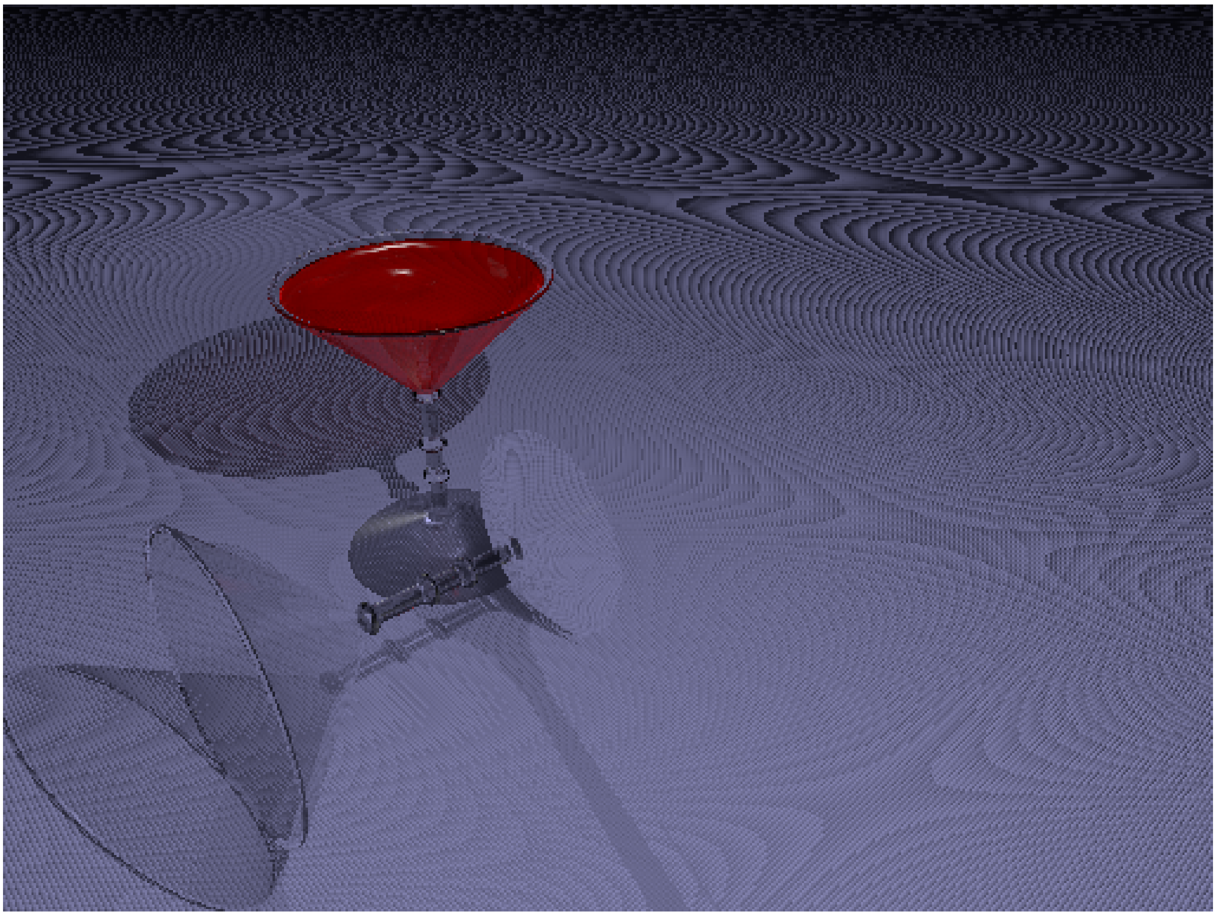

Figure 10: *Sample 2*

This scene demonstrates the various possibilities as to how photon mapping can be applied depending on the selected context regarding scene lighting environment and positioning. Different textures and object properties have significant effects on photon behaviour, thus such a scene is interesting as a case study to explore.

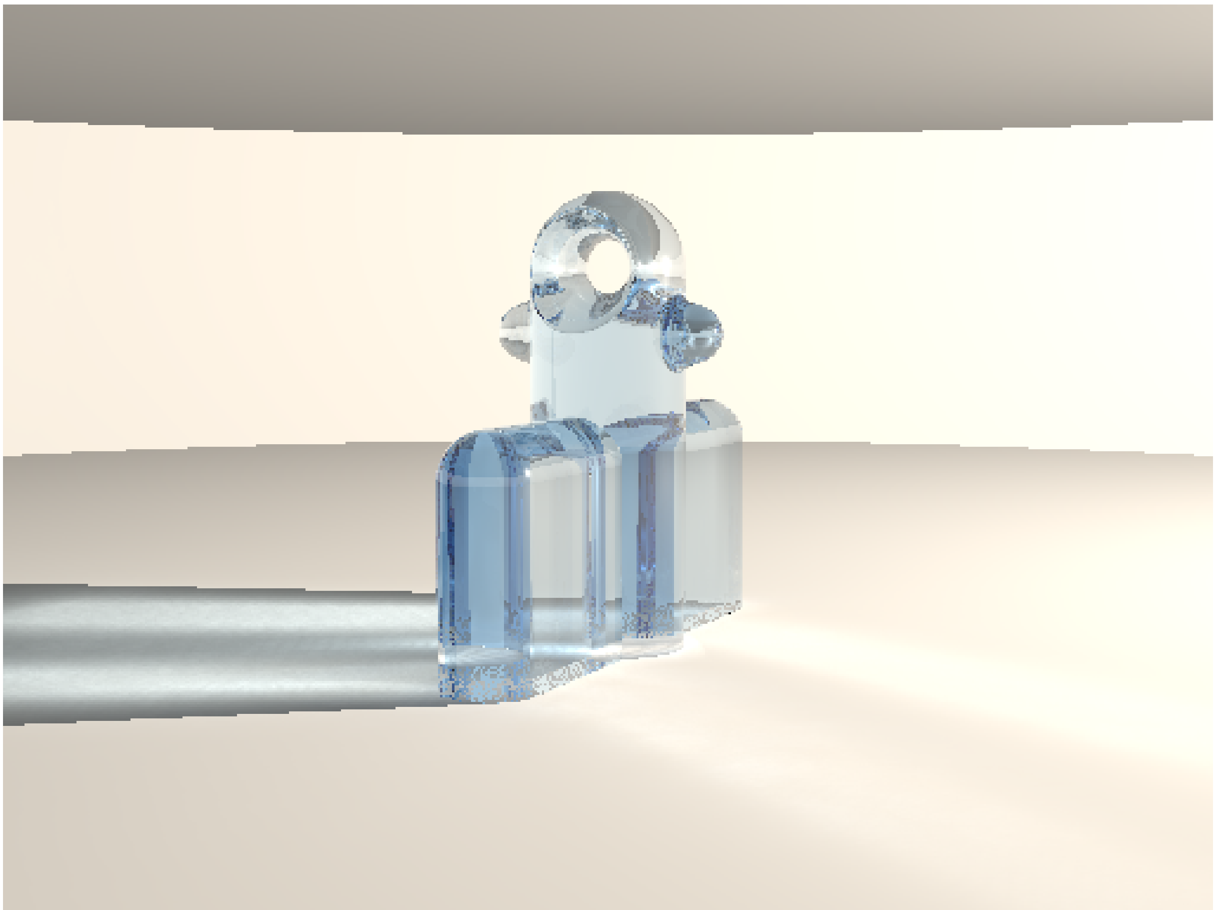

Figure 11: *Sample 3*

A simplistic scene causing caustic reflections to be espoused from the object, the layout of photon settle points is an interesting case study to experiment with regional intensity and to measure any specific rate of error when applied to K-Means clustering.

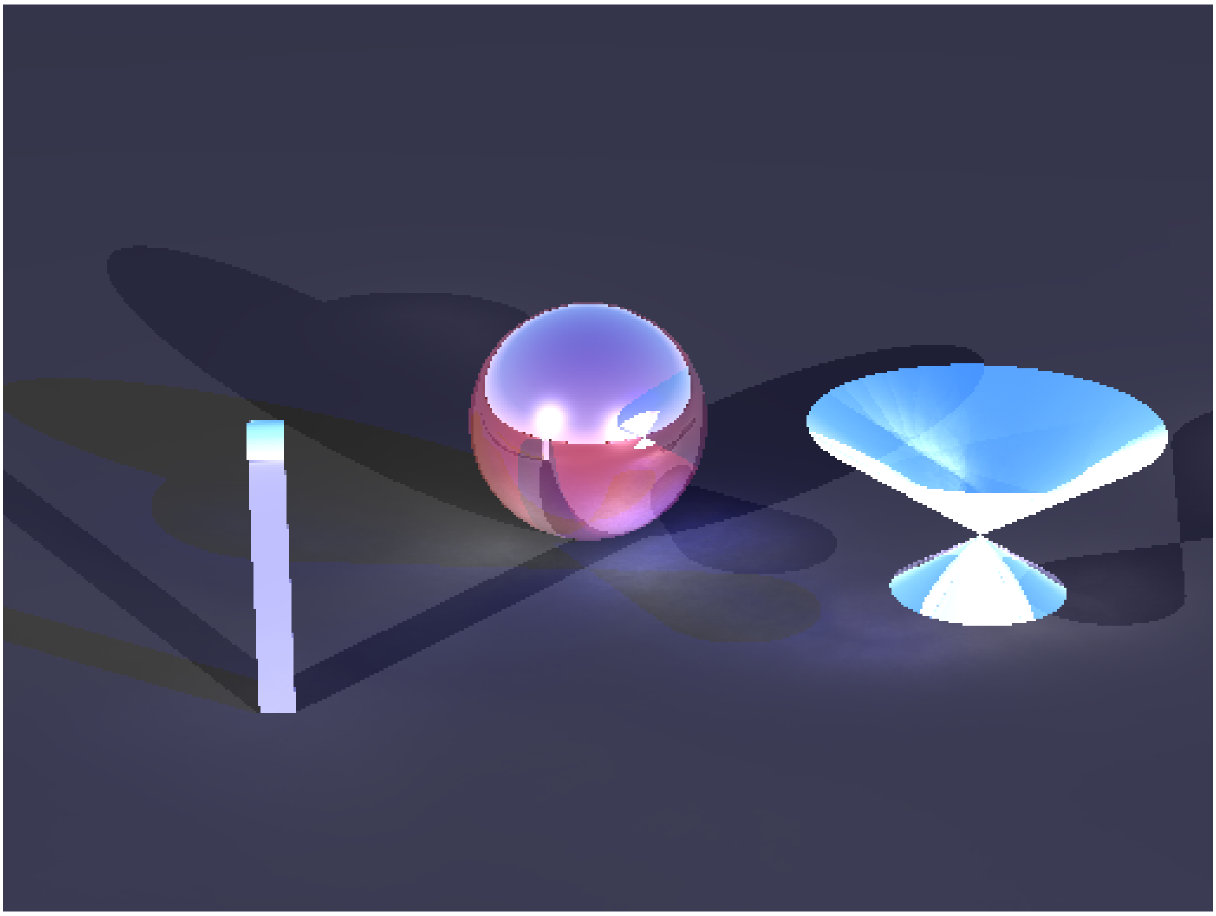

Figure 12: *Sample 4*

This scene presents an opportunity to study a variety of objects and photon collision/interactions with the scene they constitute. With multiple light sources, there also provides the opportunity to investigate the categorization properties when accommodating clustering techniques.

5.2 Querying Gather Data

We retain the knowledge that POV-Ray is program that deals largely in ray tracing and so for our photons to operate, each object must have reflective and refractive properties in order to garner sufficient photon caustic and reflective effects.

Pseudo Code

```
global_settings {
    max_trace_level 25
    photons {
        count <photons_to_shoot>
        spacing <disparity_between_photons>
        gather <min_gather>, <max_gather>
        radius <gather_radius>,<multiplier>,
                <gather_radius_media>,<multiplier>
    }
}
```

Gather Data

The POV-Ray language allows us to control in incremental steps how many photons are gathered at specific regions. This allows a steady control of input to collect from our scene which allows us to further direct our algorithm with information that corroborates with the scene at a steady pace.

The keyword gather allows us to specify how many photons are gathered at each point during the regular rendering step. The minimum number ensures a level of consistency throughout all photon populated regions, while the maximum number to gather ensures a limit cap in the case of certain regions becoming disproportionate.

Radiance Estimation

The number of photons gathered can further be defined with POV-Ray's radiance estimator. The radiance estimator is customised using a multiplier that designates the size of the target radius.

This is done with the key word radius. The larger the radius, the longer it takes to gather photons. But too small of a radius, implies the possibility of not get enough photons to get a good estimate. Normally POV-Ray looks through the photon map and uses some ad-hoc statistical analysis to determine a reasonable radius. However there is often more accuracy when overriding POV-Ray's guess and tailoring the radiance estimation for the specific scene/environment.

6 Constructing the K-Means Algorithm

6.1 Initial Abstract Approach

We present the abstract version of a K-Means clustering algorithm that can be customised for subsequent environment applications.

As with any machine learning algorithm we develop clusters of data and in this case the centre of that cluster is the centroid that respective data samples are latched to. Each data sample is assigned to a respective cluster and in regards to our application each cluster and its contextual task can be rearranged based on the situation. It is very helpful to have clusters based purely on specific regions, however clusters may also be composed based on the originating light sources of the photons.

As the objective of our investigation is to explore density estimation in a photon mapping context, successive iterations are applied through our algorithm as with traditional hierarchical K-Means clustering; for each photon latched to a centroid, assign it to that respective cluster and then we subsequently adjust the centroid according to the mean distance of the photons assigned to it.

Thus on our repeat passes we intend to recompute the centroid values. In order to do this we take all instances of a cluster, and subsequently calculate the average of them which will provide a new position for that specific centroid, this is applies to all centroids.

This iteration continues, typically until movement of the centroids is minimal. The purpose for this is accuracy, some photons will be designated to a more relative cluster than the one originally assigned.

When constructing our algorithm initially, we do it with an arbitrary data set rather than data acquired from rendering scenes.

```
function x = class5(N)

    a = 5 * [randn(N,1) + 5, randn(N,1) + 5];
    b = 5 * [randn(N,1) + 5, randn(N,1) - 5];
    c = 5 * [randn(N,1) - 5, randn(N,1) + 5];
    d = 5 * [randn(N,1) - 5, randn(N,1) - 5];
    e = 5 * [randn(N,1), randn(N,1)];

    x = [a;b;c;d;e];
end
```

This for example will provide us with 5 sets of arbitrary 'N' data of which after successive iterations of measuring the average of distance with 5 centroids, final clusters will eventually develop and settle.

It is this data set that will have to be customised when applied to rendered scenes.

6.2 Alternation of Data Through Successive Iterations

Next we plot our data in a K-Means cluster format ready for successive iterations in order to arrive at more accurate clusters of data.

First we calculate the euclidean distance between a centroid and each of its data, determining the closest data samples to each centroid.

```
while W

    dist = zeros(high,k);
    for var = 1:high
        for c = 1:k
            sumNum = 0;
            for s = 1:size(x,2)
                sumNum = sumNum + (x(var,s) - centre(c,s)).^2;
            end
            dist(var,c) = sqrt(sumNum);
        end
    end

    old_class = class;
```

We then subsequently represent the changes derived from the calculation with visual clues such as changing each data sample to the new colour depicting its relation to its respective centroid, as well as a line connecting each centroid to its sample.

```
    for var = 1:high
        [~,index] = min(dist(var,:));
         class(var) = index;
    end

    colours = rand(k,3);
    clf;

    n = length(class);
    figure(1);
    hold on;

    for i = 1 : n
        p = plot(x(i,1), x(i,2),'mx','LineWidth',3);
        set(p, 'color', colours(class(i),:))
    end

    pause(1);
    plotCentroid(centre);
    plot_lines(x,class,centre);
    pause(1);

    new_class = class;
```

At the next step we assign the newly calculated values, thus adjusting the coordinates of the centroids so that they are closer to their respective data samples in regards to the euclidean distance.

```
    for c = 1:k
        y = x(class == c,:);
        centre(c,:) = sum(y)/size(y,1);
    end

    if old_class == new_class
        W = 0;
    end

    if var_cond > max_it
        W = 0;
    end

    var_cond = var_cond + 1;
end
```

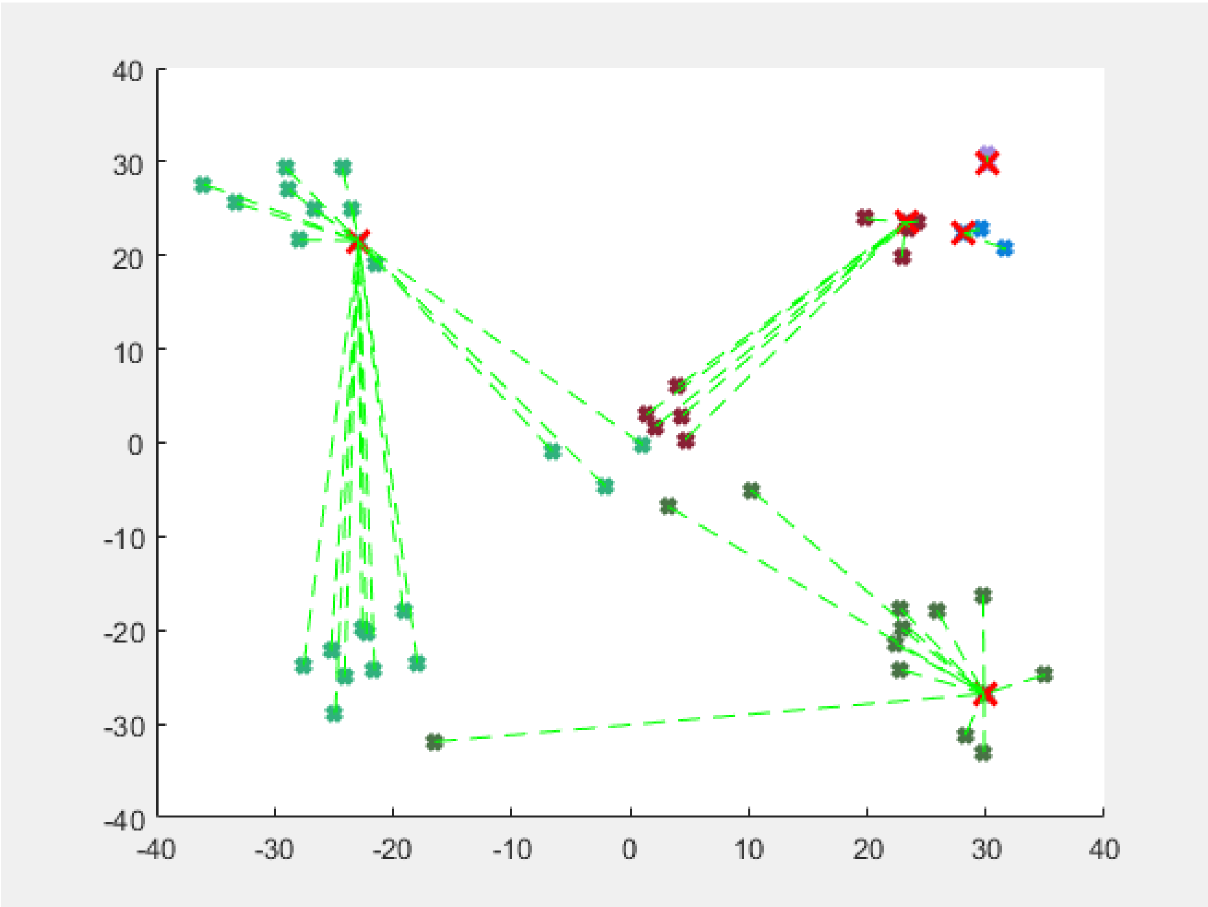

Figure 13: *Result of algorithm demonstrating initial data and centroid placement followed by euclidean calculation.*

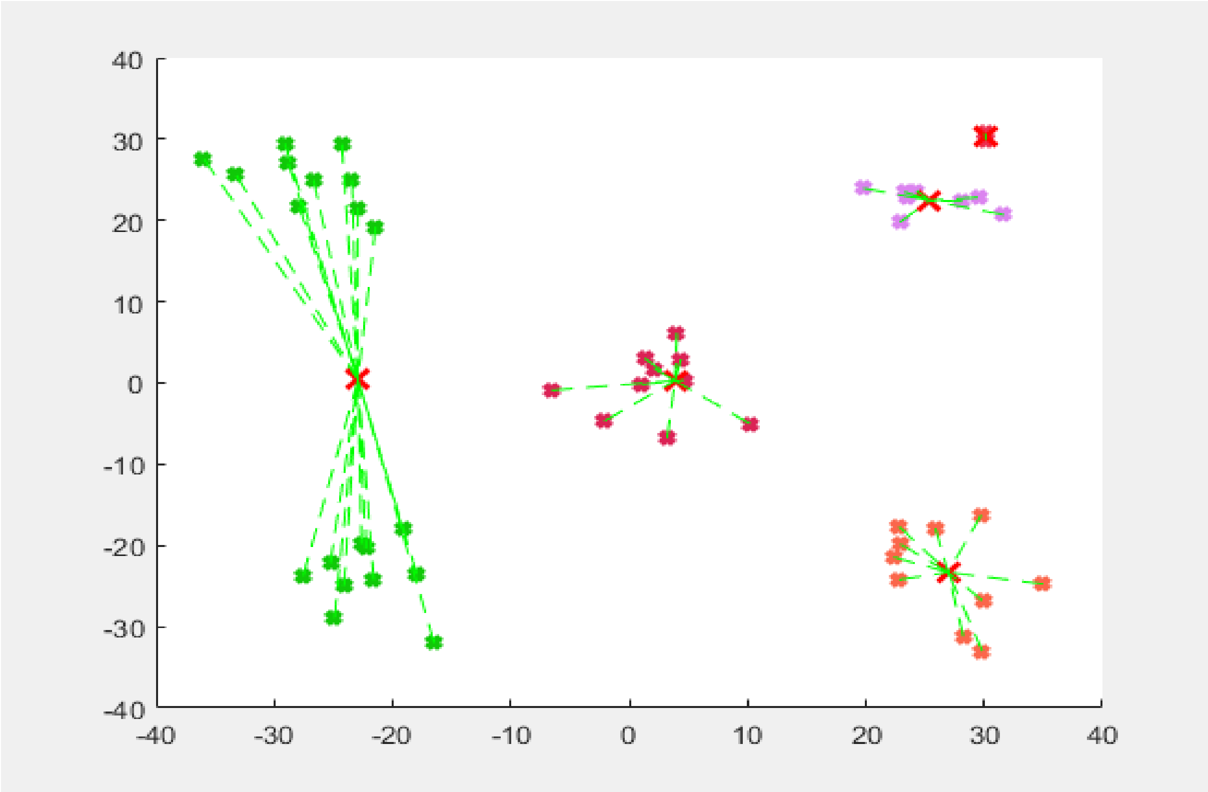

Figure 14: *Result of algorithm demonstrating further improvement of accurate placements based on successive calculations.*

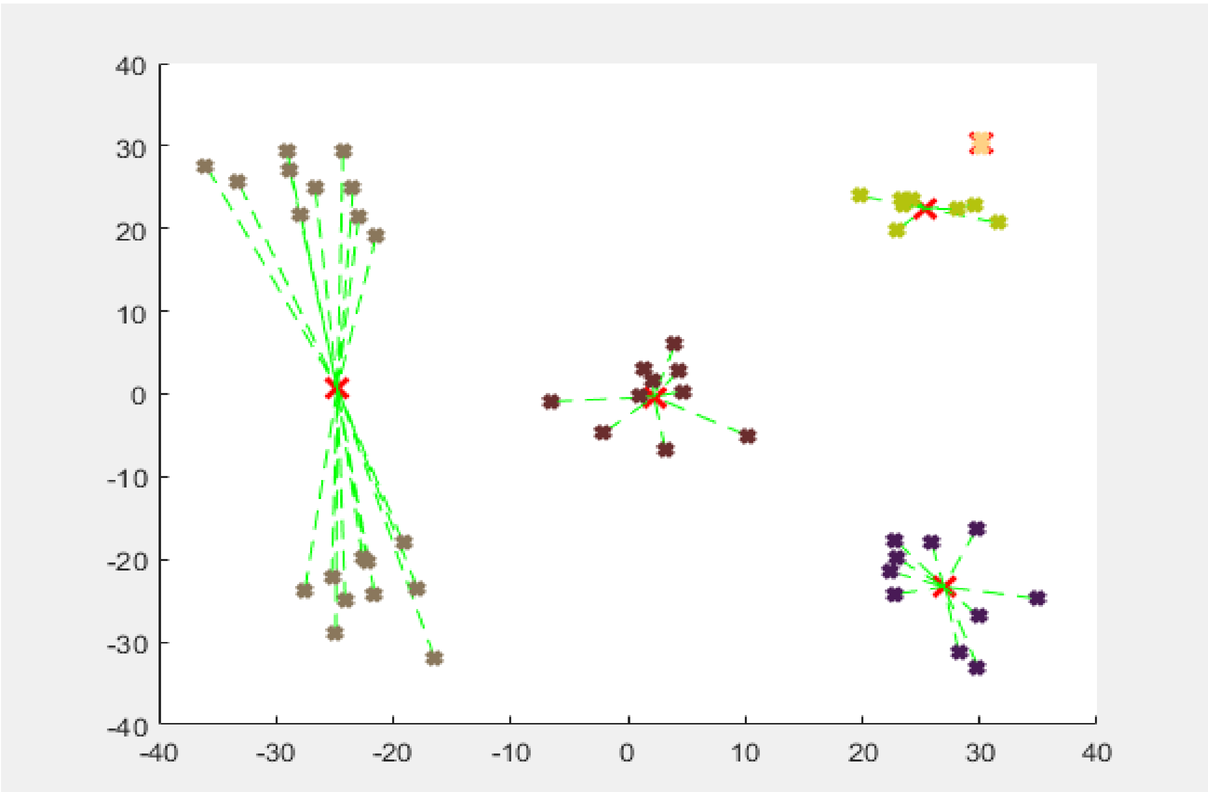

Figure 15: *Result of algorithm demonstrating finalised settle points.*

6.3 Applying Photon Map Statistics to Algorithm

Our objective is to apply this algorithm to contextual scenarios depicted in our rendered scenes.

Using gather and radiance estimator settings, we apply contextual data to our algorithm and thus in effect have K-Means operate on statistics of our environment.

In effect we are editing purely the data set to accommodate the new scenario, and this data set will be constructed using the maximum number of photons, the gather settings and the radiance estimation, it likely follows that in regards to the data we input to our algorithm there will be minor degree of error in mimicking data exactly from our rendered environments, this error could fall within the category of data sample coordinates or inaccurate gather/radiance parameters.

We retain the knowledge that manual handling and alteration of our algorithm will be a fundamental necessity of our investigation, using intuition to explore the correct number of clusters to match the amount of data samples for instance.

Since our goal is the verification of photon density estimation experiments related to our methodology requires some degree of improvisation and compromise, swapping photon samples from one region for another, thus adjusting clusters to a different and more appropriate environment for experimentation.

The K-Means clustering algorithm will provide results regarding photon samples for the purpose of evaluation, thus on the rendering side there will also be the need for manual sampling of data based on intuition, the methods of which has been discussed previously however we further insist that our investigation goes hand in hand with monitoring select gather and radiance statistics.

7 Evaluation of Methodology

Our investigation proves to yield an interesting approach to the photon mapping - density estimation procedure, from applying photon information from our rendered scenes, we see a largely abstract consensus of density estimation that requires further study, mainly in the form of more concrete practice scenarios. We may extract and emulate data samples from our rendered scenes however the effect of this on regional luminosity requires further validation.

We affirm that the K-Means clustering technique is a useful learning algorithm to decipher data samples into more accurate classifications, whereas more effort needs to be put forward to examine its effects on regional luminosity/intensity at closer inspection.

From applying a select number of test cases we explore that the K-Means algorithm has the capacity to deliver a sorting function to arrange and assort samples into a more coherent context for the efforts of global illumination effects including density estimation for the sake of accurate regional light intensity as well as the possibility to determine coherent photon relaxation.

Our investigation alludes to the possibility that our sampling strategy of relying solely on gather and radiance settings from the rendering language may not provide the most accurate assortment of data samples and hence there is room for improvement in this regard. Additionally in regards to our strategy of querying our photon map there are a number of methodologies and stratagems that may be applied.

For instance there is the possibility of including K-D trees that are to be superimposed on our rendered scene in order to better collect and assort photon samples with more control such as in Spencer and Jone's paper. This may help assist our algorithm in collecting data that is more akin to the reality of the contextual scenario as presented in the scene with more accuracy instead of relying solely on more abstract and arbitrary settings.

Additionally it is worth noting other potential areas of improvement such as working to better our algorithm by having it collect input from a scene directly and systematically rather than have the programmer manually enter data, thus crafting a more systematic procedure that can be applied at will to any scene or pre loaded photon map.

In summary the algorithm presented proves valuable in the efforts of assorting photon samples into more refined clusters so as to accurately depict regional sample proximity. K-Means proves to be a worthy tool for assisting in global illumination efforts with further refinement and practice.

8 Conclusion

Our investigation has taught us many things about the efficacy and the effectiveness of the K-Means algorithm and its unique characteristic in assorting data into more accurate classifications and categories. When applied to density estimation, practical results are still required in order to verify the authenticity of the algorithm in various contextual scenarios.

This paper has demonstrated great contextual knowledge surrounding topics dedicated to global illumination. Aside from discussing initial methodology, our investigation into this intricate topic has yielded a greater understanding of the necessities that come with advancing approaches made in photon mapping. Our investigation into this topic will only continue with time particularly as the demand for more authentic and accurate illumination effects arises and we concur that machine learning techniques and algorithms such as K-Means continues to be of greater appeal with such applications.

Overall this paper may serve as one contribution of many in the efforts of studying a refined and systematic procedure for density estimation in a graphical environment, and it is with great anticipation that we seek further testing and results with innovation in regards to collecting query data and subsequently implementing K-Means.

References

[AV06] David Arthur and Sergei Vassilvitskii. “How Slow is the K-Means Method?” In: *Proceedings of the twenty-second annual symposium on Computational geometry - SCG ’06* (2006), pp. 144–153. DOI: http://dx.doi.org/10.1145/1137856.1137880.

[BWS03] C Benthin, I Wald, and P Slasallek. “A Scalable Approach to Global Illumination”. In: *EUROGRAPHICS 2003* 22.3 (2003).

[BAJ08] B.C Budge, J.C Anderson, and K.I Joy. “Caustic Forecasting: Unbiased Estimation of Caustic Lighting for Global Illumination”. In: *Computer Graphics Forum* 27.7 (2008), pp. 1963–1970. DOI: http://dx.doi.org/10.1111/j.1467-8659.2008.01345.x.

[CWY11] J Chen, B Wang, and J.H Yong. “Improved Stochastic Progressive Photon Mapping with Metropolis Sampling”. In: *Computer Graphics Forum* 30.4 (2011), pp. 1205–1213. DOI: http://dx.doi.org/10.1111/j.1467-8659.2011.01979.x.

[DKL10] H Dammertz, A Keller, and H.P.A Lensch. “Progressive Point-Light-Based Global Illumination”. In: *Computer Graphics Forum* 29.8 (2010), pp. 2504–2515. DOI: http://dx.doi.org/10.1111/j.1467-8659.2010.01786.x.

[DJM12] Ian C. Doidge, Mark W. Jones, and Benjamin Mora. “Mixing Monte Carlo and Progressive Rendering for Improved Global Illumination”. In: *The Visual Computer* 28.6 (2012), pp. 603–612. DOI: http://dx.doi.org/10.1007/s00371-012-0703-2.

[For65] E. Forgy. “Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification”. In: *Biometrics* 21.3 (1965), pp. 768–769.

[Gau+05] Pascal Gautron et al. “Radiance Cache Splatting: A GPU Freindly Global Illumination Algorithm”. In: *ACM SIGGRAPH 2005 Sketches on - SIGGRAPH ’05* (2005). DOI: http://dx.doi.org/10.1145/1187112.1187154.

[HJ09] T Hachisuka and H.W Jensen. “Stochastic Progressive Photon Mapping”. In: *ACM Trans. Graph* 28.5 (2009), 141:1–141:8. DOI: http://dx.doi.org/10.1145/1618452.1618487.

[HOJ08] T Hachisuka, S Ogaki, and H.W Jensen. “Progressive Photon Mapping”. In: *ACM Trans. Graph* 27.5 (2008), 130:1–130:8. DOI: http://dx.doi.org/10.1145/1409060.1409083.

[HW79] J. A. Hartigan and M. A. Wong. “Algorithm AS 136: A K-Means Clustering Algorithm”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1 (1979), pp. 100–108.

[Jen96] H.W Jensen. “Global Illumination Using Photon Maps”. In: *Proceedings of the Eurographics Workshop on Rendering Techniques ’96* (1996), pp. 21–30. DOI: http://dx.doi.org/10.1007/978-3-7091-7484-5_3.

[Kel+] Alexander Keller et al. “A Divide-and-Conquer Algorithm for Simultaneous Photon Map Queries”. In: ().

[LW93] E.P Lafortune and Y.D Willems. “Bi-Directional Path Tracing”. In: *Proceedings of the International Conference on Computational Graphics and Visualization Techniques ’93* (1993), pp. 145–153.

[Llo82] S. P. Lloyd. “Least Squares Quantization in PCM”. In: *IEEE TRANSACTIONS ON INFORMATION THEORY* 28.2 (1982), pp. 129–137.

[Mac67] J. MacQueen. “Some Methods for Classification and Analysis of Multivariate Observations”. In: *Proceedings of the Fifth Berkeley Symposium On Mathematical Statistics and Probabilities* 1 (1967), pp. 281–297.

[MC13] L Morissette and S Chartier. “The k-means clustering technique: General considerations and implementation in Mathematica”. In: *Tutorials in Quantitative Methods for Psychology* (2013), pp. 15–24.

[Mys97] K Myszkowski. “Lighting Reconstruction Using Fast and Adaptive Density Estimation Techniques”. In: *Proceedings of the Eurographics Workshop on Rendering Techniques ’97* (1997), pp. 251–262. DOI: http://dx.doi.org/10.1007/978-3-7091-6858-5_23.

[Sch03] R Schregle. “Bias Compensation for Photon Maps”. In: *Computer Graphic Forum* 22.4 (2003), pp. 83–92. DOI: http://dx.doi.org/10.1111/j.1467-8659.2003.00720.x.

[Shi+95] P Shirley et al. “Global Illumination via Density-Estimation”. In: *Proceedings of 6th Workshp on Rendering* (1995), pp. 219–230. DOI: http://dx.doi.org/10.1007/978-3-7091-9430-0_21.

[SJ09a] B Spencer and M.W Jones. “Heirarchical Photon Mapping”. In: *IEEE Transactions on Visualization and Computer Graphics* 15.1 (2009), pp. 49–61.

[SJ09b] B Spencer and M.W Jones. “Into the Blue: Better Caustics Through Photon Relaxations”. In: *Eurograhics 2009, Computer Graphic Forum* 28.2 (2009), pp. 319–328. DOI: http://dx.doi.org/10.1111/j.1467-8659.2009.01371.x.

[SJ13a] B Spencer and M.W Jones. “Photon Parameterisation for Robust Relaxation Constraints”. In: *Eurograhics 2013, Computer Graphic Forum* 32.2 (2013), pp. 83–92. DOI: http://dx.doi.org/10.1111/cgf.12028.

[SJ13b] B Spencer and M.W Jones. “Progressive Photon Relaxation”. In: *ACM Trans. Graph* 32.1 (2013), pp. 1–11. DOI: http://dx.doi.org/10.1145/2421636.2421643.

[VG95] E Veach and L.J Guibas. “Optimally Combining Sampling Techniques for Monte Carlo Rendering”. In: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques ,95* (1995), pp. 419–428. DOI: http://dx.doi.org/10.1145/218380.218498.

[Viv+07] Vincent Vivanloc et al. “Light Octree: Global Illumination Fast Reconstruction and Real Time Navigation”. In: *Journal of WSCG* 15.1-3 (2007), pp. 41–50.

[Vor07] G. Voronoi. “Nouvelles applications des parametres continus a la theorie des formes quadratiques.” In: *Journal fur die Reine und Angewandte Mathematik.* 133 (1907), pp. 97–178.

[Wal98] B Walter. “Density Estimation Techniques for Global Illumination - phD Thesis”. In: (1998).

[Wan+09] R Wang et al. “An Efficient GPU Based Approach for Interactive Global Illumination”. In: *ACM Trans. Graph* 28.3 (2009), pp. 1–8. DOI: http://dx.doi.org/10.1145/1531326.1531397.