Selecting Island Pixels With a Likelihood Ratio

Jordan Tweddale

6/24/2013
Senior Project

California Polytechnic State University
Department of Physics

Supervisor: Dr. Jodi Christiansen

Abstract:

A new method, called the Log Likelihood Ratio (LLR), was employed in place of a Signal-to-Noise Ratio (SNR) to pick out signal pixels on the VERITAS (Very Energetic Radiation Imaging Telescope Array System) telescopes. The LLR test was performed on both photoelectrons and timing for VERITAS to select island pixels in Stage 2 cleaning, which is where the camera image is processed to determine which pixels on the camera are signal, and which are background. The two LLRs were combined to form a final LLR, which has a numerical cut, LLR = 11.5, between signal and background pixels. The combined LLR was found to be a powerful indicator of what pixels are or are not signal pixels.

Table of Contents

I. Introduction

Telescope cameras typically use statistical methods to distinguish light from a particular source and light from night-sky backgrounds and electronic noise. All telescopes rely on the information provided by light, so it is therefore very important for a telescope to have a high resolution to provide reliable data. My goal is to employ a new method for selecting signal pixels for VERITAS. Employing this new technique has the potential to increase the resolution of the images produced by VERITAS telescopes and also lower the energy threshold of the instrument, giving VERITAS better results. First of all, increased resolution allows for a more accurate measure of where the signal is coming from in the sky. The lowered threshold has a similar effect. This allows the shape and size of the image produced by the telescope to be better understood. Since the direction and size of the source are the main things that astronomy relies on, the proposed LLR technique is expected to improve the significance of the measurements made by VERITAS.

II. The Science of Veritas

VERITAS stands for Very Energetic Radiation Imaging Telescope Array System. VERITAS is a system of four 12 m Cherenkov telescopes located at the Fred Lawrence Whipple Observatory in southern Arizona[1]. These telescopes are used to observe gamma-rays in the very high energy (VHE) range of 50 GeV – 50 TeV. There are more than 100 recorded sources detected in the VHE regime. The presence of high-energy cosmic gamma-rays usually implies extreme conditions, such as high magnetic or electric fields, or stellar explosions, which accelerate particles to relativistic speeds. Because of this, VHE gamma-rays are more closely associated with supernovae, pulsars, and blazars, rather than stars or normal galaxies[2], which produce light through black-body radiation.

Supernovae are the remnants of an exploded star. For example, the Crab nebula (also called M1) is one of the most famous supernovae, and has a strong source of VHE gamma-rays extending up to the TeV range. This supernova explosion occurred about 1000 years ago, on July 4, 1054, and was first discovered to be in the VHE range in 1989 at the Whipple observatory[1]. Pulsars are highly magnetized, rotating neutron stars; the most famous pulsar in the VHE range is Geminga, which is located about 200 parsecs away from the Earth. Blazars, however, are the most common source of the VHE regime. These are a type of active galaxy thought to have a supermassive black hole at their center with a relativistic jet coming out and pointing in the direction of Earth. These are slightly different than quasars, which are viewed at an angle away from the jet. The difference between these types of galaxies with relativistic jets is illustrated in Figure 1. Since most gamma-rays come from the jet, Blazars (which look directly down the jet) are the most common source of VHE gamma-rays, Quasars (which look at an angle to the jet) are slightly less common sources, and Radio Galaxies (which look at a 90^{o} angle to the jet) are VHE quiet.

There is a fundamental problem in observing the VHE range; gamma-rays do not penetrate the Earth's atmosphere. At lower energies (less than about 30 GeV) gamma-rays are detected from instruments attached to satellites or balloons, which have a collection area the same size as the instrument (around $1m^2$) . However, in the VHE range, a much larger collection area is needed, because the gamma-ray flux is too small to be observed from space. This problem is solved using the Cherenkov Technique. Particles incident on the Earth's atmosphere trigger a particle cascade, also called an air shower. Charged particles that pass through a dielectric medium, such as the atmosphere of the Earth, with a speed greater than the phase velocity of light produces electromagnetic radiation in the direction of the air shower. This radiation is called Cherenkov radiation, named after Russian scientist Pavel Alekseyevich Cherenkov, who discovered the effect in 1934 and in turn won the 1958 Nobel Prize for his discovery.

Typical air showers yield a light pool on the ground with a radius of about 130m, with the maximum emission (for 0.1 – 1 TeV radiation) occurring at 10km above the Earth's surface. The photons in the air shower arrive in a brief flash, called a "Cherenkov Pulse", which lasts for a few nanoseconds. This pulse is detected anywhere in the light pool by using reflective surfaces to focus the light onto a detector. Because of this, Atmospheric Cherenkov detectors (as they are called) can achieve collection areas of about 10^5 m^2 [2] . VERITAS uses an array of four 12 m telescopes made up of 350 large segmented mirrors. The telescopes are placed on a steerable altitude-azimuth mount and each has a field of view of about 3.5 degrees[1]. The basic principles behind the Cherenkov technique are shown in Figure 2.

A camera consisting of 499 photomultiplier tubes is placed at the focus of the mirror, each PMT is considered a pixel on the camera. PMTs are highly sensitive light detectors that multiply the current incident on the tube by several orders of magnitude. Each PMT's photon count is combined with the readings from other PMTs in the camera to record the total "shape" of the shower. As shown in Figure 2, gamma-ray events, which come from the sources mentioned before, typically have the shape of a long ellipse. The long axis of the ellipse corresponds to the vertical extension of the shower and points toward the position of the source. In the case of VERITAS, the shapes from three additional telescopes are used to do a stereo reconstruction of the incident gamma or cosmic ray. The position of the source is the intersection point of the various image axes.

Currently, VERITAS uses the signal-to-noise ratio (SNR) to determine which pixels on the telescope camera are called signal, and those which are regarded as background. The SNR is simply the ratio of the number of photoelectrons (PEs) counted by a photomultiplier tube (PMT) divided by the electronic jitter, called the pedestal variance. The higher the SNR, the more likely it is that the pixel has meaningful information. The SNR is commonly used in astronomy, imaging, and electrical engineering. Even though the SNR is widely accepted and has provided good results, a new method, called the Log-Likelihood Ratio (LLR), is being introduced. The LLR is statistically more sensitive than the simpler SNR method, because it compares how much more likely it is that a given model is true over another.

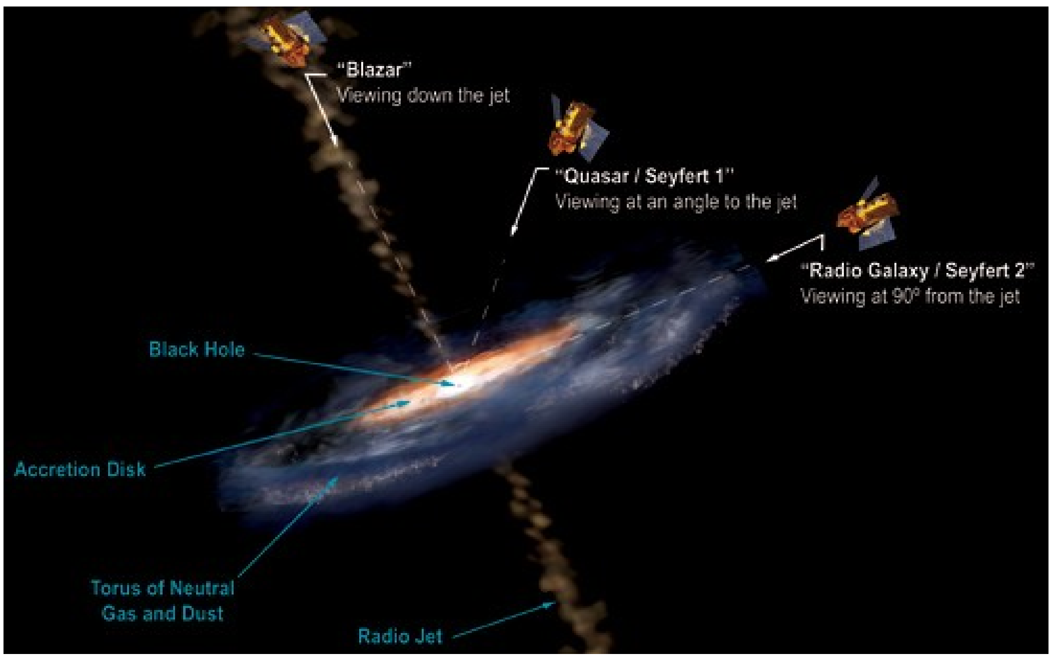

Figure 1[3] : Active galaxies (AGN) with supermassive black holes and relativistic jets have been classified into a variety of names based on the viewing angle. The classifications predate the understanding we now have of an AGN.

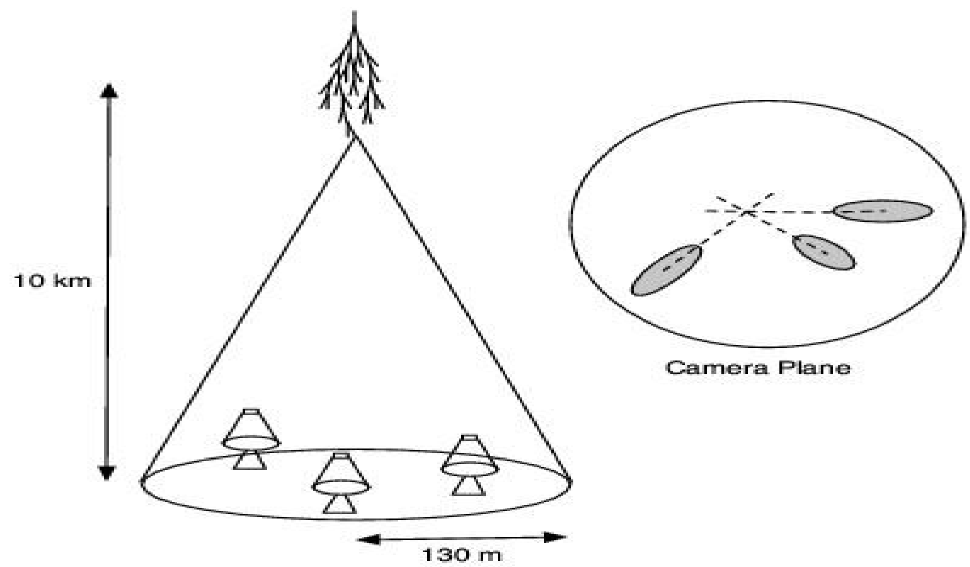

Figure 2 [2]**:** The telescopes are positioned inside the light pool, the images from each telescope are combined to determine where the shower is coming from.

III. Selecting Pixels to Analyze

Currently in the VERITAS event display cleaning process, called stage 2, any pixel on a telescope whose SNR is above 5 is called signal. An SNR of 5 simply means that the measured number of photoelectrons on a pixel (signal) is at least five times greater than the pedestal variance (noise)[4]. A pixel with an SNR between 2.5 and 5 is called signal only if it neighbors a pixel with an SNR greater than 5. Any pixel not satisfying these conditions are background pixels and are not considered to constitute part of the image of the air shower. Camera plots of the SNR are shown in Figure 3 for a Crab Nebula event, and in Figure 4 for a simulated gamma-ray event. The red island of pixels in each event is evidence of light.

The Log-Likelihood Ratio is a little more complex than the signal-to-noise ratio. The basic definition of the LLR is given by:

$$D = -2 \ln \left(\frac{\text{Likelihood for Null Model}}{\text{Likelihood for Signal Model}} \right) \tag{1}$$

The likelihood for the null model and the signal model are both defined by the Gaussian function given by:

$$P(x)=(1/\sigma\sqrt{(2\pi)})e^{(-(x-\mu)^2/2\sigma^2)} \tag{2}$$

where σ is the standard deviation, x is the number of PEs on the pixel, and μ is the mean number of PEs for the model. Among two models where there are no unknown parameters, such as the case we're working in here, the LLR has the highest statistical power among its competitors[5].

The null model assumes that the number of PEs on all of the pixels is due to background noise. In this case, μ is zero, and σ is given by the pedestal variance (a statistical measure of the electronic jitter on background pixels). The negative exponent causes the value of this Gaussian to disappear when the difference between x and μ is large, which means the pixel has a very high number of PEs. The converse is also true. The value of the Gaussian reaches its maximum when the value of x is close to zero. This means that the null model is large for pixels that are background and is small otherwise.

The signal model assumes that all of the counts collected by the pixels are due to the signal, in this case a VHE gamma-ray. For this model, x is defined the same as it is for the null model (number of PEs). However, now μ is given by the average number of counts, we estimate μ using the signal in each pixel's neighboring pixels. Again, σ is the standard deviation, but this time it's given by the combined uncertainty in the counts and the pedestal variance:

$$\sigma=\sqrt{(\mu+pv^2)} \tag{3}$$

where pv is the pedestal variance. Now, the Gaussian has very small values when the counts in one pixel are very different from the counts in its neighbors. There is a little more going on here though. If a pixel's count is close to zero and its neighbors are all close to zero, then shouldn't it be picked out as signal in this model? After all, that was the case for the null model. The truth is, this model merely picks out which pixels have values that are close to the values of their neighboring pixels.

Remember however, that this value for the signal model is going back inside Equation 1. The ratio of the likelihoods is close to a value of 1, which makes the logarithm tend towards zero. The ratio of likelihoods for a pixel with a high number of PEs and a value close to that of its neighbors will be very small, so Equation 1 will give a high value for D, meaning that this is probably a signal pixel. Conversely, if a pixel has a high number of PEs but differs from its neighbors, then Equation 1 will give an intermediate value for D, which will be relatively high or very low depending on how much the pixel differs from its neighbors.

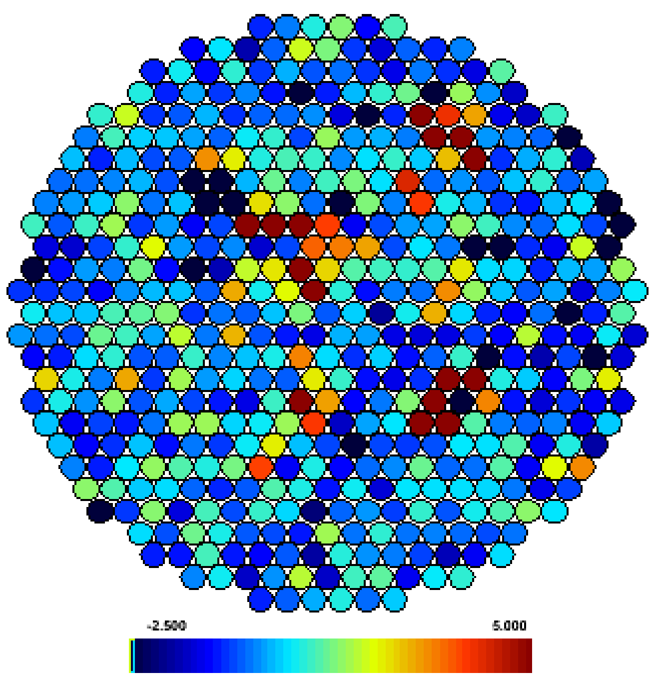

Figure 3: SNR plot of an event from the Crab Nebula. Pixels that are red have an SNR greater than 5 and are considered signal, neighboring pixels that have SNR above 2.5 (orange) are also called signal. The dark blue pixels are either dead pixels on the telescope or are blacked out due to the presence of a star in that area of the sky.

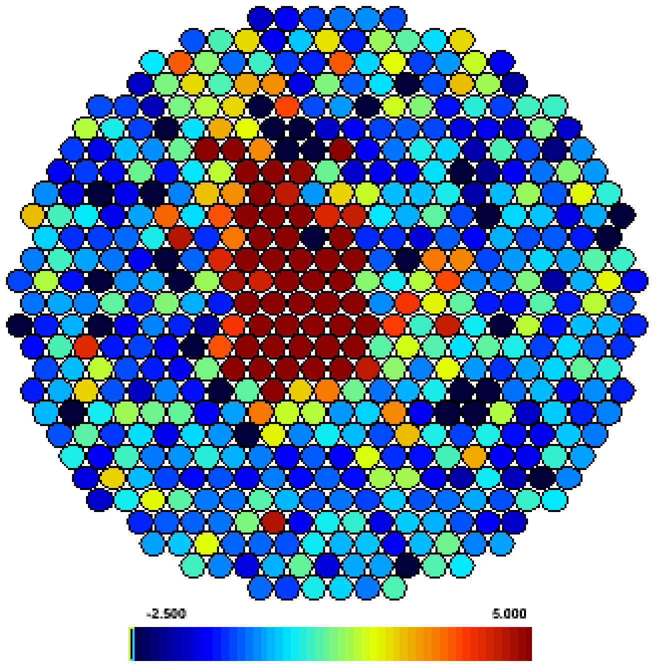

Figure 4: SNR plot of a simulated event.

IV. Optimal PE LLR Cut

The next thing that needs to be decided is the value of D in Equation 1 that dictates the cutoff between signal and background. Since the LLR will be replacing the SNR in the VERITAS code base, it should have at least the same level of accuracy in determining signal as the SNR. To do this, we found the number of times the SNR cut of 5 accidentally selected a pixel for events triggered without any light. In a single Crab file, VERITAS has about 210,000 total events, and 798 of these are pedestal events. Pedestal events are simply a measurement of the photoelectrons on the camera when there is no signal present. Theoretically, the pedestal events should take the form of a Gaussian the same width as the pedestal variance, centered around zero. However, every once in a while, there will be a few pixels in the tail of the Gaussian that are far from zero, most likely from low energy cosmic muons passing close by the telescope. These outlying pixels are used to determine how often the SNR cut accidentally selects a pixel, and in turn, where the cut for the LLR should be to achieve the same level of accidents.

Over these 798 pedestal events, there are a total of 68 pixels that have an SNR above 5. Remember that each event has nearly 500 pixels, this means that the SNR cut is wrong a little less than 0.017% of the time. Statistically, a significance of 5σ (SNR = 5) should only be wrong 0.000057% of the time[4], meaning that over these 798 events, there should only be 0.22 false positive pixels. The reason there are so many more false positive pixels than expected is because moderate energy muons are striking the telescope and polluting the pedestal events. A histogram of the SNR for these 798 pedestal events is shown in Figure 5. A histogram of the LLR for the 798 pedestal events is shown in Figure 6. The LLR was run over the pedestal events and it was determined that 68 pixels lied outside a value of 9.7. This is where the cut between background and signal is placed.

With the cut between background and signal established, the LLR was run over the remaining array events (~209,000) in the Crab file. Histograms of the SNR and the LLR of each pixel in the array triggered events are shown in Figures 7 and 8 respectively. Note that these histograms are similar to their corresponding pedestal histograms in Figures 5 and 6, except with longer tails indicating signal pixels. Also, note the spike on the LLR plot at 0 and the longer tail on the SNR plot. This means that the background for the LLR appears to be more definitely background, and the signal appears to be more definitely signal. For a visual representation of this effect, refer to Figures 9 and 10, which are camera plots of the SNR and LLR for the same event. These two figures show that the background pixels are better defined for the LLR than the

SNR, meaning that the LLR is less noisy.

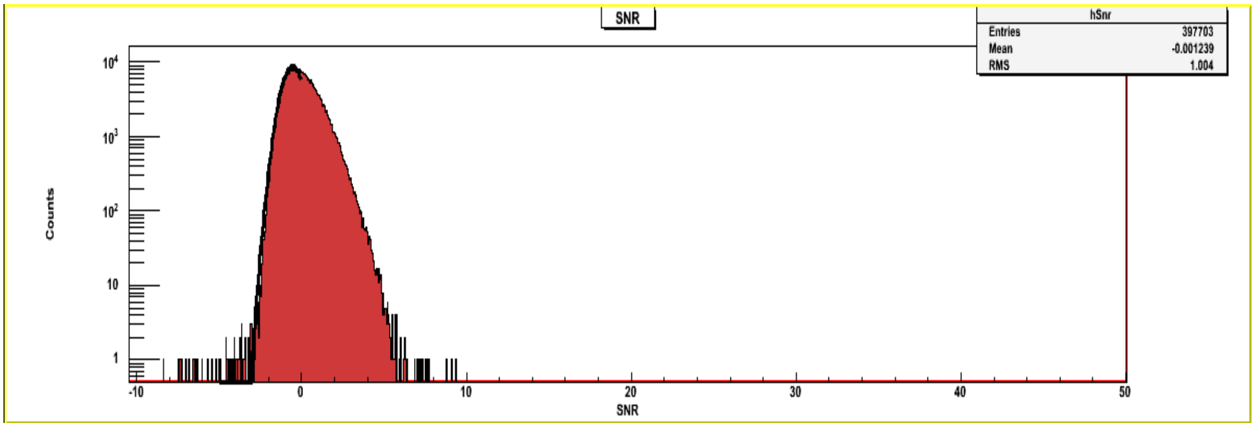

Figure 5: Histogram of SNR for all pixels in 798 pedestal events.

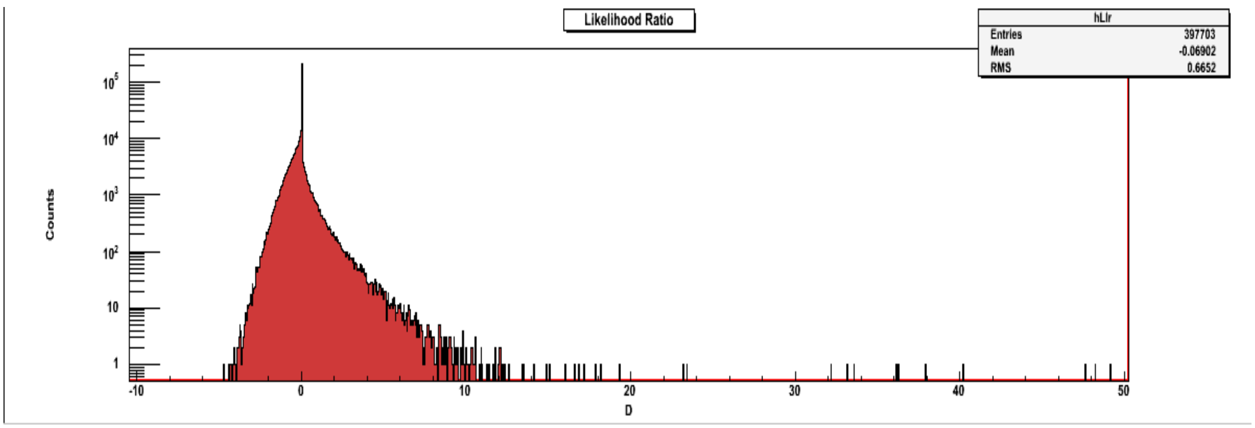

Figure 6: Histogram of LLR for all pixels in 798 pedestal events (Note: there are two more pixels outside D = 50).

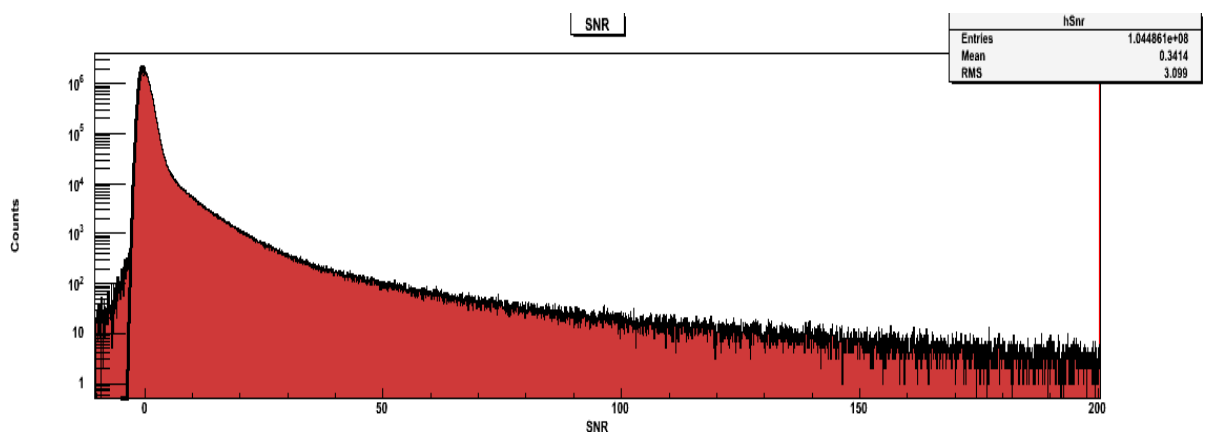

Figure 7: Histogram of SNR for all pixels in array triggered events (Note: there are pixels with values above SNR = 200).

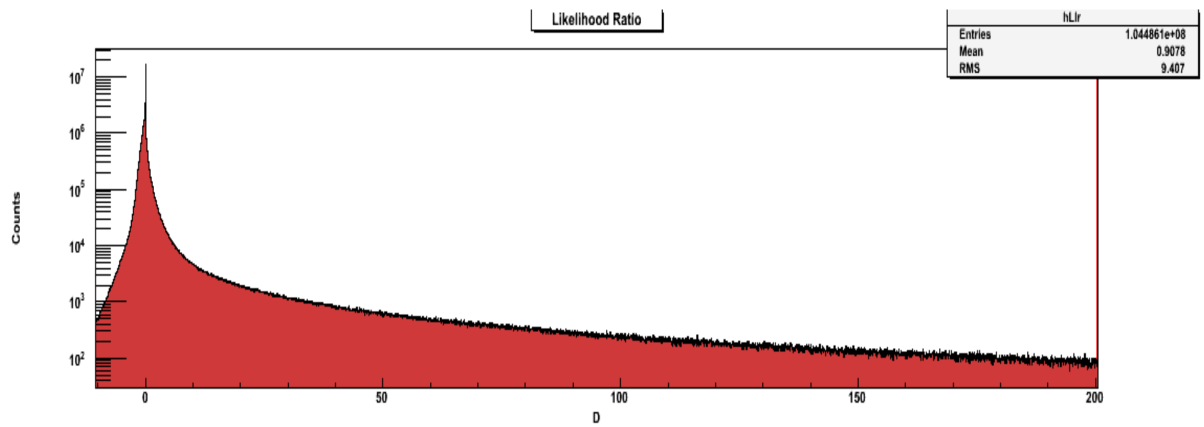

Figure 8: Histogram of LLR for all pixels in array triggered events (Note: there are pixels with values above D = 200).

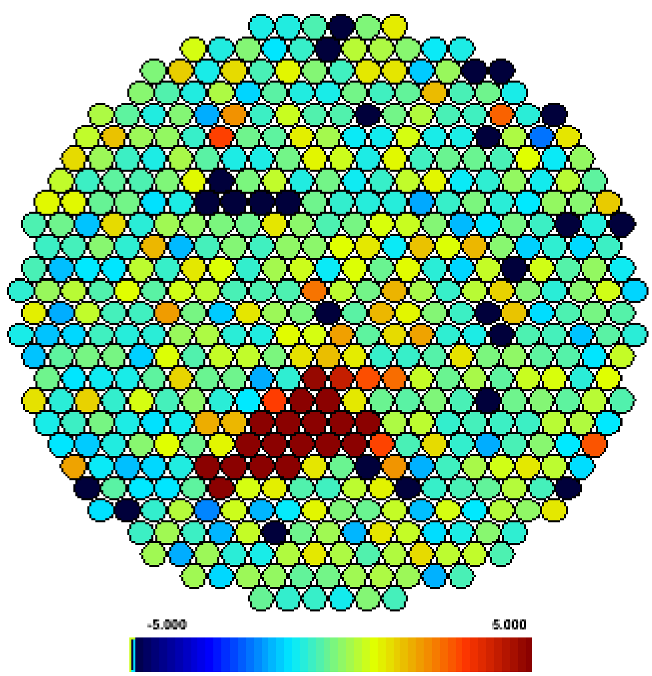

Figure 9: SNR plot of an event from the Crab Nebula. Pixels that are red have an SNR greater than 5 and are considered signal. The dark blue pixels are either dead pixels on the telescope or are blacked out due to the presence of a star in that area of the sky.

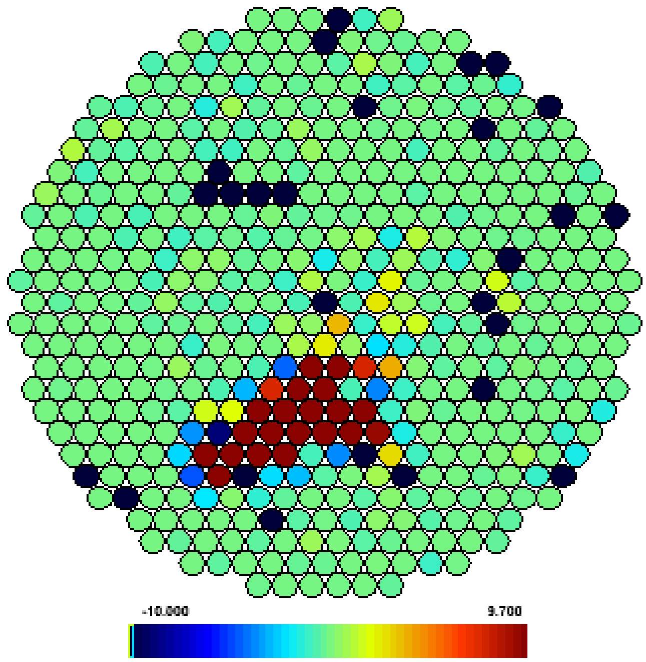

Figure 10: LLR plot of an event (same event as Figure 9) from the Crab Nebula. Pixels that are red have an LLR greater than 9.7 and are considered signal. The dark blue pixels are either dead pixels on the telescope or are blacked out due to the presence of a star in that area of the sky.

V. Timing Probabilities

The next step is to introduce a new LLR with timing probabilities in addition to the LLR from the PE probabilities. Each event has a variable, called t_0, which is defined as the time at which each pixel reaches half of its maximum voltage. A pixel that has counts due to signal has a t_0 similar to other pixels which have counts due to signal, this is because the air shower arrives at these pixels and cause a max voltage at nearly the same time for each of these pixels. On the other hand, the maximum voltage for background pixels, which is due to the electronic jitter of the PMT, can come at any time. Examples of what the voltage vs. time plots might look like for an event are shown in Figure 11 (for signal) and Figure 12 (for background). A camera plot of the

t_0 on each pixel is shown in Figure 13. Notice that the patch of signal events all have a time that is nearly identical.

To create the timing LLR, a new variable was introduced, called Δt_0. The Δt_0 for a pixel is defined as the average t_0 of a pixels neighbors minus the pixel's t_0. The average t_0 is weighted by the number of PEs in the pixel, so neighboring pixels that have a higher PE count will be given more weight towards the average t_0. The basic idea here is similar to that of the photoelectron counts, if a pixel has a t_0 close to that of its neighbors, then that pixel is likely a signal pixel (since the air shower will cause pixels to trigger a voltage spike at the same time). Background pixels, on the other hand, will probably have a t_0 much different than their neighbors. To sum this up, if a pixel has a Δt_0 that is close to zero, it is probably a signal pixel, and if it has a Δt_0 that is either much higher or lower than zero, then the pixel probably isn't a signal pixel. A camera plot of each pixel's Δt_0 is shown in Figure 14.

To make an LLR from Δt_0, we had to look at all of the array and pedestal events again. A histogram of the Δt_0 for pedestal events is shown in Figure 15, while a histogram of the Δt_0 for array triggered events with a photoelectron LLR greater than 9.7 (thus signal pixels) is shown in Figure 16. Now refer back to Equation 1. The likelihood for the null model and the likelihood for the signal model are found using the histograms in Figures 15 and 16. The likelihood for the null model and signal model is given by the number of events in a particular Δt_0 bin divided by the total number events. For the null model, the pedestal event histogram is used (Figure 15). The signal model uses the array triggered histogram (Figure 16). This gives a timing LLR which is based solely on a pixel's Δt_0 value. A plot of the timing LLR vs. Δt_0 is shown in Figure 17. Note that pixels that have a Δt_0 value close to zero are assigned a positive LLR value (with a maximum LLR of about 5), and those that are farther away from zero are assigned a negative LLR value. This means that a pixel with a Δt_0 close to zero is likely a signal pixel based on these timing statistics. A camera plot of the timing LLR is shown in Figure 18. Compare this to Figure 19, which is the PE LLR for the same event (duplicate of Figure 10). The Timing LLR is clearly noisier than the PE LLR, but still contains some statistical power to detect signal.

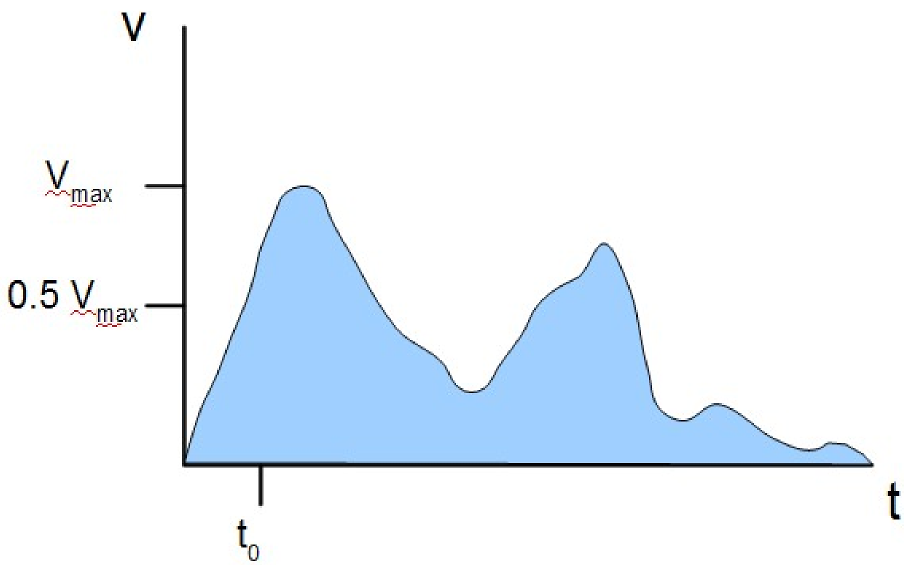

Figure 11: Sketch of a voltage vs. time plot for a shower triggered pixel. t_0 is defined as the time when the voltage reaches half of its maximum, signal pixels will have a similar t_0. The area under the curve is how the PE count is defined for a pixel

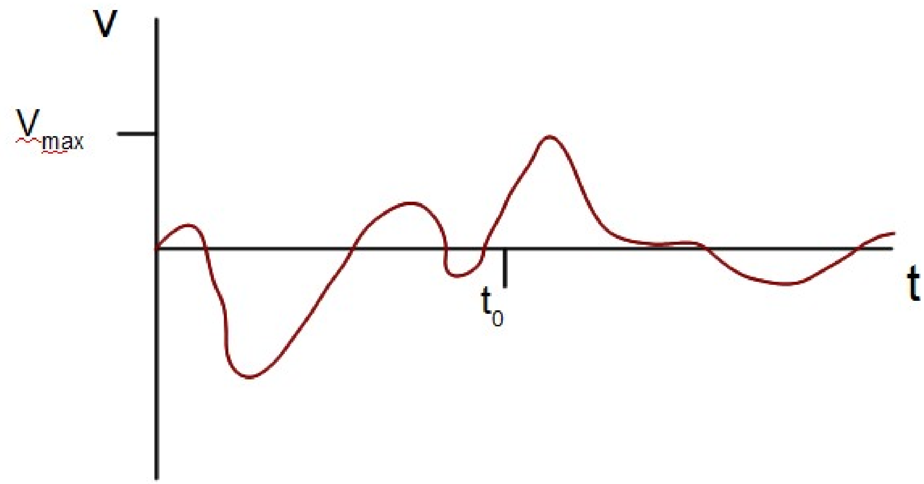

Figure 12: Sketch of a typical voltage vs. time plot for a background pixel. t_0 is defined as the time when the voltage reaches half of its maximum, background pixels will have a dissimilar t_0.

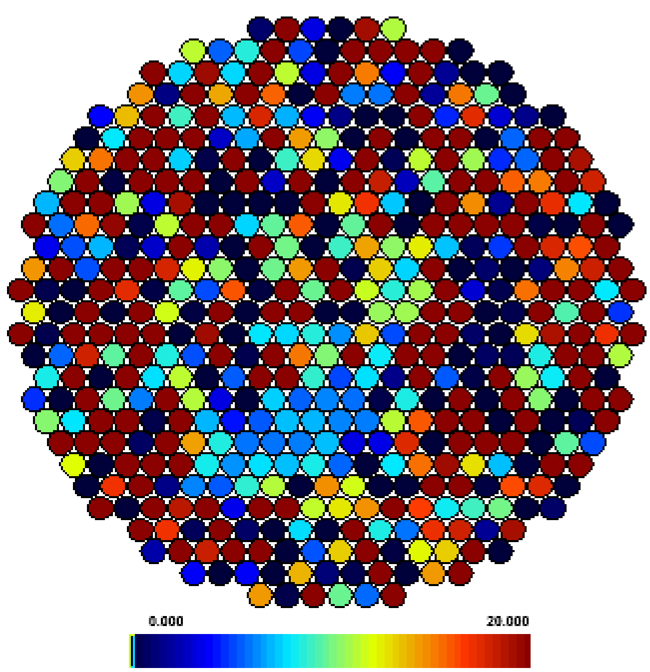

Figure 13: Camera plot of t_0 for a single event (same event as Figures 9 and 10), the signal pixels all have a similar t_0 while the background pixels can have almost any t_0

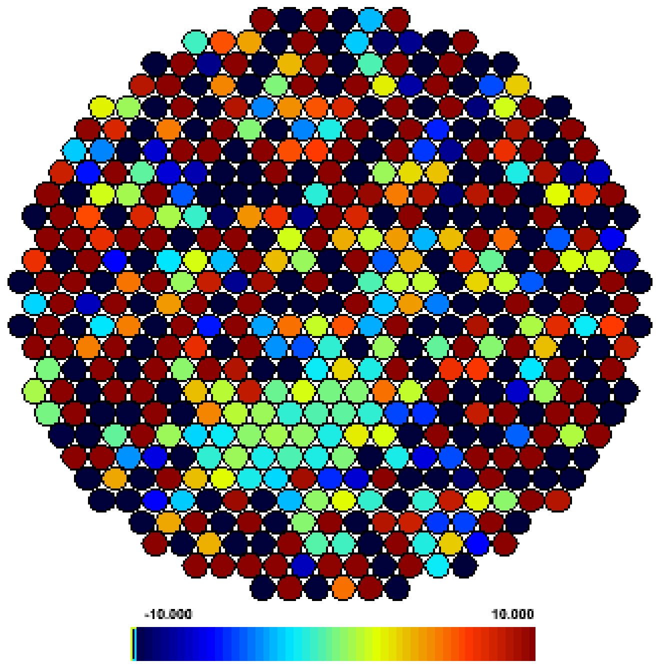

Figure 14: Camera plot of Δt_0. Pixels that are green have a Δt_0 close to zero and are probably due to signal.

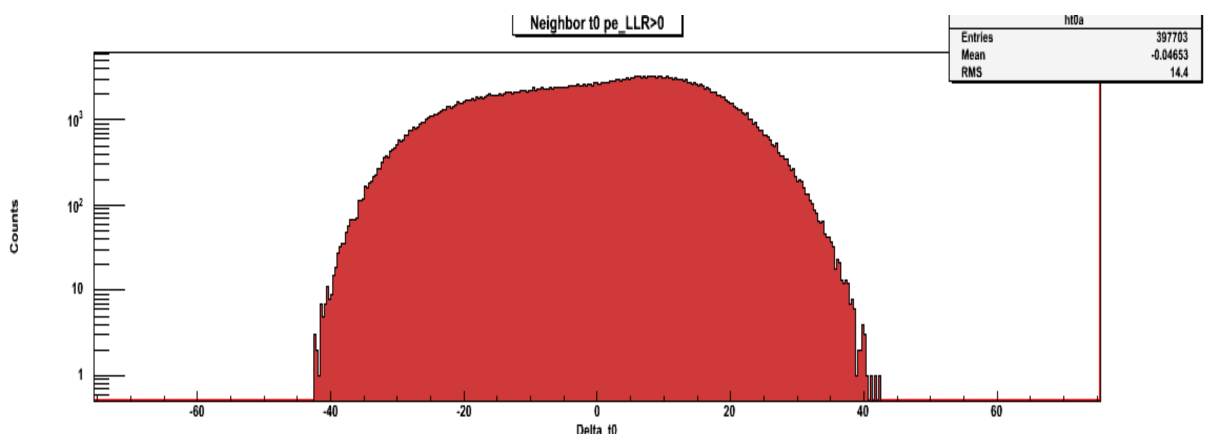

Figure 15: Histogram of Δt_0 for pedestal events, note that Δt_0 is nearly equally likely in between -20 and +20.

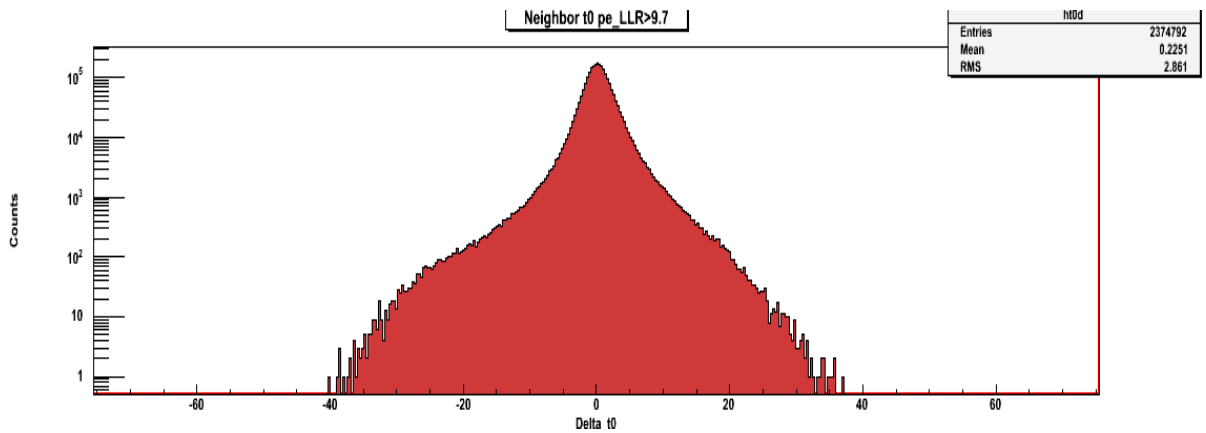

Figure 16: Histogram of Δt_0 for array triggered events with a PE LLR > 9.7. Note the spike around $\Delta t_0 = 0$ is due to the presence of signal.

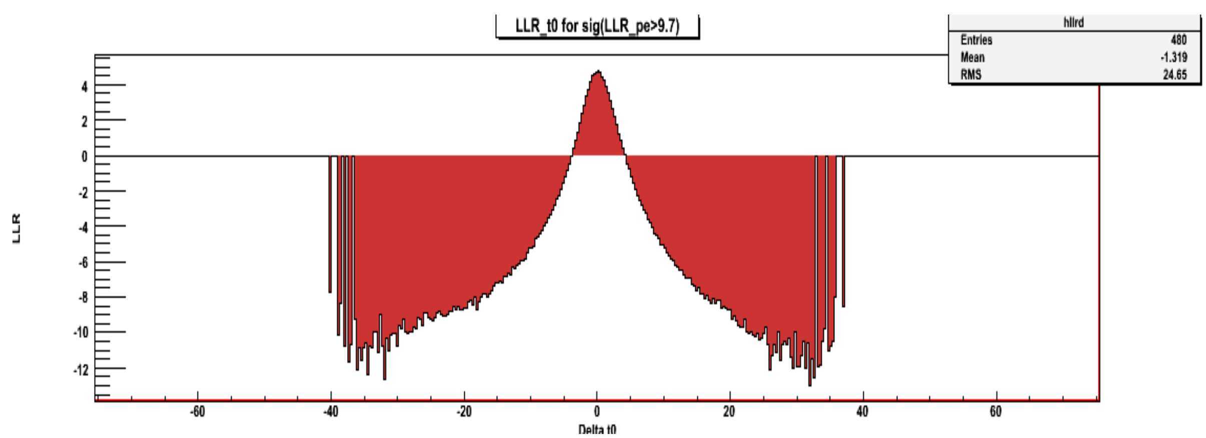

Figure 17: Timing LLR vs. Δt_0. Pixels with a Δt_0 value relatively close to zero have a positive LLR (meaning that they're more likely signal) while pixels whose Δt_0 is farther away from zero have a negative LLR (meaning that they're more likely background). Note that the timing LLR has a maximum value of about 5.

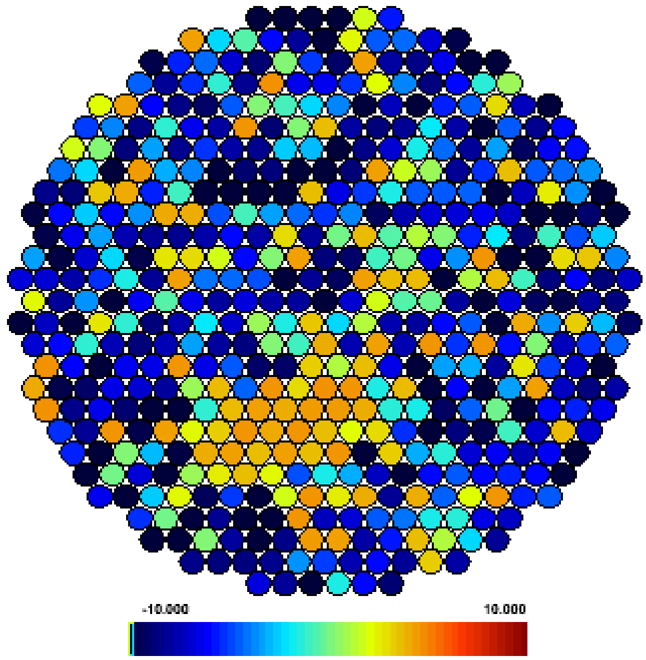

Figure 18: Camera plot of timing LLR for a single event. Orange pixels have an LLR at or close to the maximum value of about 5, and are more likely to be signal.

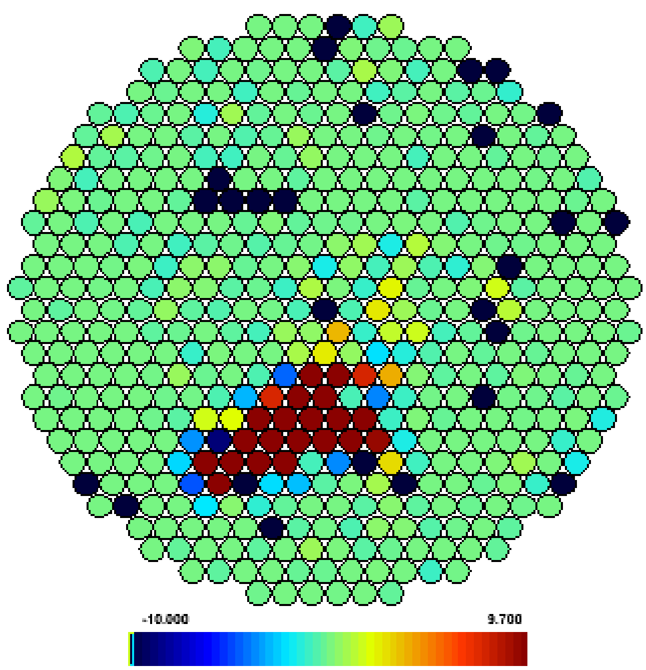

Figure 19: Camera plot of PE LLR for the same event as Figure 18.

VI. Combined LLR

The final step is to combine both the timing LLR with the PE LLR to make one final combined likelihood ratio. When the two LLRs are combined, the final combined LLR looks like:

$$D = -2\ln\left(\frac{P_{null}(\Delta t_0)\ P_{null}(PE)}{P_{signal}(\Delta t_0)\ P_{signal}(PE)}\right)$$

which is just:

$$D = -2\ln\left(\frac{P_{null}(\Delta t_0)}{P_{signal}(\Delta t_0)}\right) - 2\ln\left(\frac{P_{null}(PE)}{P_{signal}(PE)}\right)$$

and can be reduced to:

$$D = LLR(\Delta t_0) + LLR(PE) \qquad (4)$$

The two LLRs are simply added together to make the combined LLR. A new signal cut was needed for this combined LLR, so the same process was applied to the combined LLR as was applied to the PE LLR. This time, the corresponding 68 pixel cut for the pedestal histogram (Figure 20) was at a value of D = 11.5. The histogram for the combined LLR for array triggered events is shown in Figure 21. This combined LLR is significant because it slightly raises the LLR value for pixels that are in time with their neighbors, and slightly decreases the LLR value for pixels who are out of time with their neighbors. This further separates background pixels from signal pixels, and also picks up a few more pixels as signal that may have just missed the cut for the PE LLR, but *did* make the cut for the combined LLR (because they were in time with their neighbors). A camera plot of the combined LLR for a single event is shown in Figure 22, compare this to Figure 23 which shows a camera plot of the SNR of the event, and Figure 24 which shows a camera plot of the PE LLR of the event. It can be seen that the combined LLR seems to be a little noisier, but the combined cut is more powerful at rejecting spurious pixels.

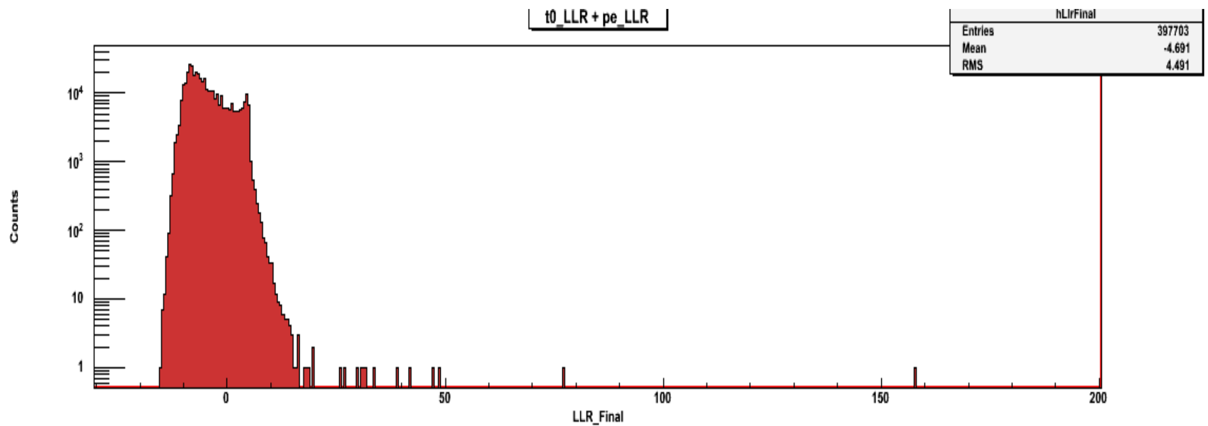

Figure 20: Histogram of the combined LLR for pedestal events, this is used to determine what the signal cut is for the combined LLR.

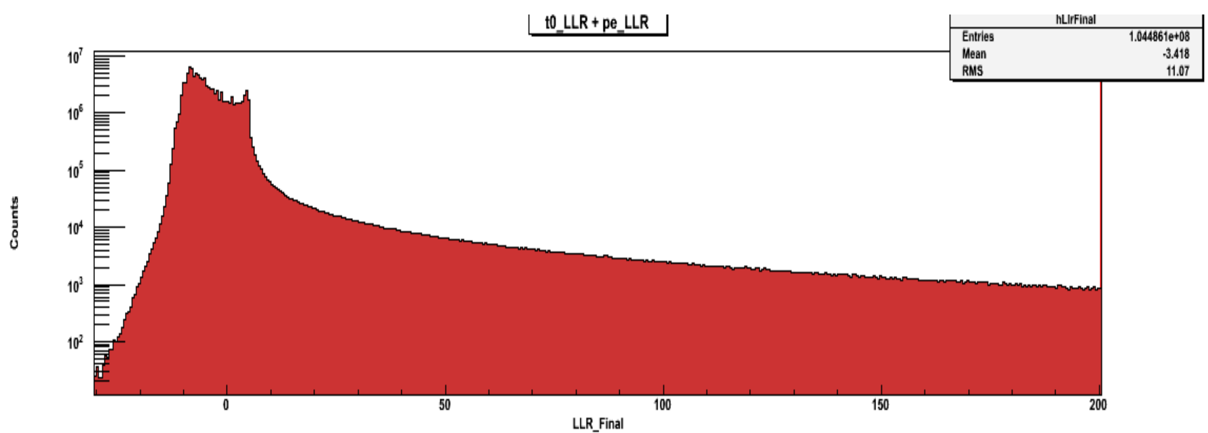

Figure 21: Histogram of the combined LLR for array triggered events.

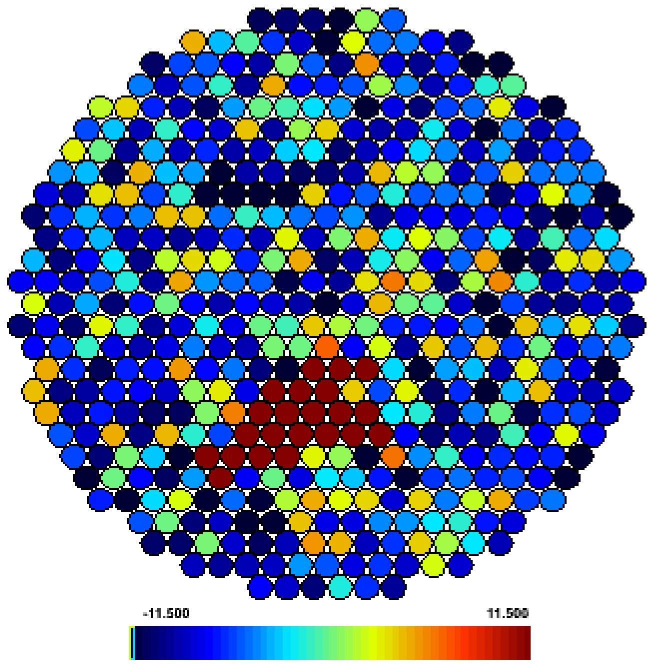

Figure 22: Camera plot of combined LLR for a single event.

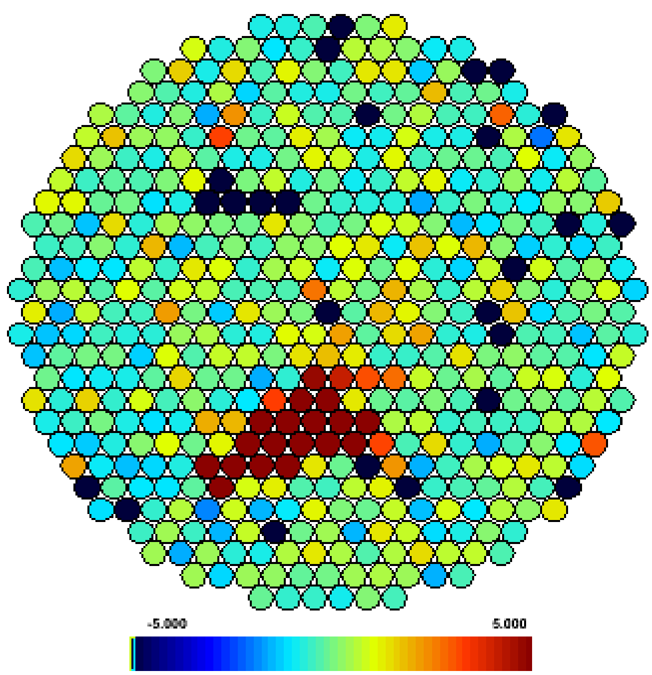

Figure 23: Camera plot of SNR of a single event.

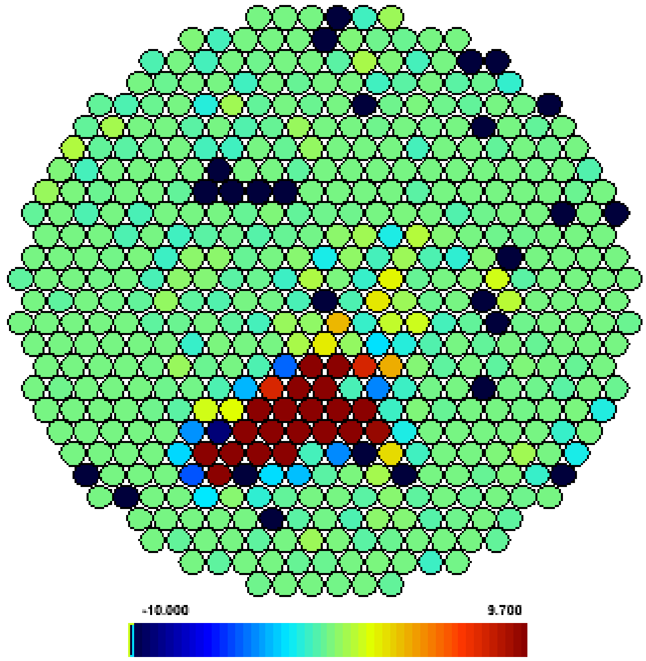

Figure 24: Camera plot of PE LLR of a single event.

VII. Conclusion

We have shown that the LLR technique clearly has the potential to improve the VERITAS resolution and sensitivity. The PE LLR combined with the timing LLR is a powerful indicator of what is or isn't signal, because it both picks up more pixels than the SNR method and is less likely to pick up pixels that aren't actually a part of the island. Another advantage that the combined LLR has over the SNR is that it avoids using the two-tier selection rules that the SNR has. This means that a pixel won't be rejected as a signal pixel based on the “signal status” of its neighbors. It doesn't matter much that the addition of the timing statistics causes the combined LLR to be noisier than the PE LLR. This is because any pixel that is a background pixel in the

PE LLR can have a maximum combined LLR of about 5 (assuming its in time with the island pixels), which significantly misses the cut of 11.5. However, a pixel on the edge of the island that just misses the PE LLR cut, but is in time with the island, will be picked up by the combined LLR. All of this can be summed up as an increase in the resolution of the VERITAS camera images, which fulfills the goal that we set out for.

VIII. References

[1] Holder, J. et al.(VERITAS collaboration) 2006. The first VERITAS telescope. *Astroparticle Physics*, 25 (6) , 391-401.

[2] http://veritas.sao.arizona.edu

[3] http://www.nasa.gov/centers/goddard/news/topstory/2007/active_galaxy.html

[4] Taylor, J.R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements* (2nd Ed). University Science Books.

[5] Neyman, J. & Pearson, E.S. 1932. On the Problem of the Most Efficient Tests of Statistical Hypothesis. *Philharmonic Transactions of the Royal Society*, 231 (694-706), 289-337.