Response to Rebuttal

I am writing a response to Zhang and Yau's rebuttal to my review of their paper "Simultaneous three-dimensional geometry and color texture acquisition using a single-chip color camera".

After reading the rebuttal and the edited text, I feel there has been some improvement in stating the principles of the approach, although I feel it is still not explained coherently in a single paragraph or section. Also, I unfortunately do not feel the major issues with the paper's organization, analysis, discussion have been resolved.

While the language remains quite readable, the organization has actually become more muddled with the revision. A vast majority of the changes to the paper have been added into the introduction instead of more appropriate sections. I still find the analysis and discussion lacking, including some unexplained inconsistencies. I also have spotted a potential technical problem in their approach which affects two of their major claims.

Although I will discuss these issues in the markup of the revised text and rebuttal, I will mention the potential technical problem in the cover letter, which I believe will highlight the need for a more coherent explanation of the principles, a deeper analysis and discussion, and a revised organization of the paper. I also wish to highlight this problem here so that it isn't lost in the markup, since I feel it directly affects the claims of the paper.

Zhang and Yau claim that they are capable of simultaneously acquiring 3D geometry and texture using a single-chip camera with the acquisition being insensitive to object color effects and being capable of achieving pixel level 3D resolution. However, I believe this may not always be the case when using a Bayer pattern camera. The following example will illuminate the problem:

Suppose you have a purely red object, and you sequentially project the three BW fringes onto the scene. The camera views the scene and each Bayer cell sees the corresponding color component of the fringes modulated by the object's color. For each pixel, the phase estimate is supposed to be unaffected by the object color, since it sees only one color for all three images. But since the object is purely red, the camera pixels under the blue and green Bayer filters do not receive any light and therefore can't make a phase estimate.

As the filter colors are interspersed, the phase estimate cannot be made at the full resolution of the camera, as is claimed in the paper. In a worst case scenario, the 3D resolution is half the camera resolution, just as it is for the color texture. Smoothing the result will not provide the lost resolution, especially at edges or other discontinuities in position or slope.

This appears to be the case in the calibration target experiment, in which a 250mmx250mm red/blue checkerboard is imaged by a 640x480 camera and the phase and color are estimated. In this case, each red checkerboard cell is similar to the above example. If you look at the plot of the reconstructed phase, the resolution is only 320 pixels across. Although the axis is labeled in millimeters it is most likely in pixels, since the target is only 250mm wide but the axis goes from -50 to 270. Also, if you count the samples (which fall at integer units on the x-axis) or look at the resolution of the noise, there is appears to be a 1 sample/unit, which also implies 320 samples of resolution for a 640 pixel width camera. I would find it very hard to believe it was 640 samples across with samples falling on half-integer units of the x-axis.

Similarly, it cannot be claimed that the approach is insensitive to object color effects, as this example shows the phase estimate and resolution are indeed affected.

There was no similar analysis or discussion in the paper with regards to this.

This appears to be an implementation fault of using a Bayer pattern and not a fault of the main approach (using three BW phase shifted fringes, averaging them for color texture, and doing point-by-point phase estimation). This would not occur if, for example, a 3CCD camera were used, since for each point at least one color component would be
sensed for a non-black object.

I mention this not because I favor a 3CCD, which I agree is more expensive than a Bayer camera; but thinking of the problem in terms of using what actually makes the system work versus what is implementation brings up many questions about the applicability, limitations, and trade-offs of the system, as well as what is novel and the real contribution.

Unfortunately, this is the type of analysis and discussion that appears to be lacking in the paper, and I requested in the initial review. It also illustrates why in the organization of the paper there needs to be a clear description and distinction between what makes the approach work and what is implementation, as well as a distinction between the which advantages/disadvantages can be attributed to the approach versus the implementation. I also requested this in the initial review.

While I highly admire the end results and I'm sure many people at conferences were equally impressed and are eagerly awaiting the published paper, the standards for a peer-reviewed journal paper should be higher than a conference paper. I stand by my previous review and restate my previous recommendation: I do recommend that this paper be published after the major issues regarding the claims have been resolved, but either the paper should be converted to a Letter (which doesn't have the depth of a paper, but I don't believe gives the paper's end result the proper exposure it deserves), or additional analysis should be performed for the work to be of significance and depth for a full paper and to present the work in its best light.

Suggestions
Some specific suggestions (in order of importance) which I feel are critical:

• Resolve the object color problem and full resolution question. Either state that they are problems of using a Bayer pattern rather than problems of the main approach, or fully state and defend that isn't a problem in the Discussion section.

• Explain in one location (besides the abstract) your goals. Simultaneous/colocated camera acquisition for 3D/color, aligned 3D/color, insensitivity to color bleeding and object color, high speed, cost.

• Explain in one location, what makes the approach work (the operating principle), how it meets your goals, and how it is novel. See original review's summary.

    The color coupling is avoided because a B/W pattern is projected. The phase can be computed point by point from the three phase shifted sinusoids, and the three images can be averaged to remove the fringe information and retrieve color texture. The surface color affects each of the three B/W images the same way, and so phase is computed from three phase shifted sinusoids that have been affected by the surface color in the same way, thereby avoiding the affects of surface color.

Notice a choice of implementation isn't mentioned here (Bayer, 3CCD, foveon, etc.)

• Explain why you chose the Bayer pattern camera to implement the approach and how it is compatible with the operating principle and at what possible expense.

• In the advantages/disadvantages section explain which advantages/disadvantages are related to the operating principle and which are related to the implementation (Bayer pattern).

    Simultaneous/colocated camera acquisition for 3D/color, aligned 3D/color, insensitivity to color bleeding and object color, high speed are all operating principle advantages, not dependent on Bayer pattern implementation.
Larger local noise than BW camera, 1/2 camera 3D and color resolution, 2/3 light loss are all Bayer pattern problems not operating principle problems.

3CCD implementation has full resolution, lower light loss, less local noise, etc compared with Bayer implementation. What are the tradeoffs (cost, alignment)?

• Fix the units (mm to pixels) in the red/blue target cross section plot, explain why it isn't full resolution and how reduced, explain why the acquisition time is not 1/3 the camera frame rate (see markup).

• State advantages/disadvantages with regards to comparable systems (BW for fringe and color camera for texture), or state what system you are comparing with and why. (see markup for more guidance).

Single color camera system vs. BW cam for 3D plus color cam for texture -- single camera reduced alignment issues, cheaper, etc.

Single color camera (Bayert) versus BW camera for fringe detection (since both are must capture fringes and fringe contrast is an issue this is a valid parallel comparison) -- BW camera may have shorter exposure time (see text), single camera can capture 3D at same cost as BW with color texture for free.

Zhang/yau 3 frames vs. Zhang/tower 12 frames.

• Add more analysis and discussion more than just an obvious advantage/disadvantage list, including phase estimate insensitivity to color balance and aberrations, explanation of artifacts in the 3D data, noise, etc. I would again like to know the scanning volume and related accuracy, even just in the caption of the figure. In this way I can get a sense of if its applicable to my needs.

• Principles in principles section not in intro. Add cost goal in abstract. 3CCD vs bayer in either principle (Bayer) or discussion section. Mention in Principles section calibration is needed to go from phase to displacement, but do calibration explanation (red/blue card) in setup section. Do system cost in setup section.

REVISED DOCUMENT

Abstract
Include low cost as a goal, since you state in the rebuttal (but not in the original abstract or introduction) that is important to you.

Page 4
"In our RECENTLY efforts"
recent

"However, obtaining precisely aligned geometry and color texture REMAINS NOT VERY DIFFICULT"
remains very difficult

Page 5
"ON THE CONTRAST,"
In contrast,

Zhang, Tower, Tower 2008 "Shape and colour ..." paper didn't include the color model and nonlinear regression. It was their 2007 "Phase and color calculation..." paper.
In their 2008 paper, they sequentially project the 3 color fringes, and take the channel with the maximum modulation to determine the phase. This is how they deal with the object surface color. At a low level, it is similar to your approach in that in a sense you are projecting red, blue, green fringes (mixed to be white), and sensing them with the same color channel in your camera. So in a sense, object color problems are solved using multiple sets of monochrome fringes (either consecutively projected or simultaneously) being sensed by multiple corresponding monochrome sensors/measurements (either separate or multiplexed).

So, the below statement is not quite right because they project color fringes and are able to measure the surface profile (and texture) properly (even if the object is purely red).

In general, all the techniques using color fringe images suffer the problem: they may not measure the surface profile properly. For example, for a red object, the green and blue light will be absorbed, and the information carried on by these color fringes will be lost. Therefore, for a structured light system, to measure object’s 3-D geometry correctly, a black-and-white (B/W) camera and the white light is usually desirable.

Yes, their technique will not be as fast as yours (an advantage you should point out), since they use up to 12 projections (mainly to avoid a wrapped phase map and to uniquely identify fringe orders).

At the end of their paper, they state "it will be interesting to test composite fringe patterns... which means 3 color channels are used simultaneously, so that acquisition time will reduce to 1/3". This is essentially your method, which is something you should point out! And they probably can't do a composite pattern, since their fringes are of different frequencies to optimally detect the fringe order for absolute phase determination (versus wrapped and using an unwrapping algorithm). Therefore, they will still have color bleeding problems.

"They SUFFERS the same light loss as the single-chip color cameras since each sensor only captures partial spectra of the light."

Note:
Although Wikipedia is not exactly the best source to quote
http://en.wikipedia.org/wiki/Charge-coupled_device

Another advantage of 3CCD over a Bayer mask device is higher quantum efficiency (and therefore higher light sensitivity for a given aperture size). This is because in a 3CCD device most of the light entering the aperture is captured by a sensor, while a Bayer mask absorbs a high proportion (about 2/3) of the light falling on each CCD pixel.

This is the point I was trying to make in the review. The light isn't absorbed by a filter in a 3CCD, its redirected to a different sensor. These sensor readings can be combined to either reduce noise or even to pick the best channel to use. Whereas, in the Bayer, the light is absorbed and lost.

The alignment problem for 3CCD is acknowledged as a potential problem, but its typically aligned to subpixel accuracy at the factory and has adjustment pins to realign. Its much different than aligning a camera-projector pair with a beam splitter. (What about a Foveon, which gives 3 colocated color measurements? It is more expensive though and maybe noisy at lower light levels).

"We SIMPLIFY generate fringe images"

simply

"projected by the projector AT exactly the same manner,"
We SIMPLIFY generate fringe images
simply
(stated earlier. prefer to remove earlier sentences).

Why does the camera run at 70fps @ 640x480 (mentioned in the setup section), but data acquisition takes place in 250ms (4fps) (mentioned in the experiment section -- you need to mention the difference)? Shouldn't you get 70/3=26 fps or 38ms for the data acquisition? (>15fps would be considered real-time in many areas). Is it to avoid needing synchronization between camera and projector? If so, you should mention this. It is important to your setup, and without it, it causes confusion later in the advantages section. Do you average frames later or pick good candidate images?

Are the units really mm in the x-direction? Most likely the units are pixels, since target is only 250x250mm (stated in rebuttal), but the x-axis and image goes from -50 to 270 which implies 320 pixels. Likewise looking at the noise or counting samples, its about 1 sample/tick, or 320 samples (pixels).

Why is the horizontal cross section only 320 pixels? Shouldn't it be a full 640 pixel resolution? Is this because you are having red a blue checkers, so in every color square you are missing every other phase estimate, and missing phase estimates on every other line?

Scanning volume and accuracy really should be mentioned, as well as size and distance to target. As you can see it was useful for me to know the size and distance to interpret and judge your plot.

Less sensitive to surface color variation and High Resolution as stated may not be accurate claims

Zhang and Yau claim that they are capable of simultaneously acquiring 3D geometry and texture using a single-chip camera with the acquisition being insensitive to object color effects and being capable of achieving pixel level 3D resolution. However, I believe this may not always be the case when using a Bayer pattern camera. The following example will illuminate the problem:

Suppose you have a purely red object, and you sequentially project the three BW fringes onto the scene. The camera views the scene and each Bayer cell sees the corresponding color component of the fringes modulated by the object's color. For each pixel, the phase estimate is supposed to be unaffected by the object color, since it sees only one color for all three images. But since the object is purely red, the camera pixels under the blue and green Bayer filters do not receive any light and therefore can't make a phase estimate.

The Bayer pattern is laid out as such

RGRGRG
GBGBGB

As such, the first line will see

RxRxRx

where x is no light captured and hence phase cannot be estimated. So a phase estimate is only made every other pixel.

The second line will see

xxxxxx
So, no phase estimate can be made on this entire line!

As the filter colors are interspersed, the phase estimate cannot be made at the full resolution of the camera, as is claimed in the paper. In a worst case scenario, the 3D resolution is half the camera resolution, just as it is for the color texture. Smoothing the result will not provide the lost resolution, especially at edges or other discontinuities in position or slope.

This appears to be the case in the calibration target experiment, in which a 250mmx250mm red/blue checkerboard is imaged by a 640x480 camera and the phase and color are estimated. In this case, each red checkerboard cell is similar to the above example; likewise for the blue cells. If you look at the plot of the reconstructed phase, the resolution is only 320 pixels across. Although the axis is labeled in millimeters it is most likely in pixels, since the target is only 250mm wide but the axis goes from -50 to 270. Also, If you count the samples or look at the resolution of the noise, there is appears to be a 1 sample/tick, which also implies 320 samples of resolution for a 640 pixel width camera.

Similarly, it cannot be claimed that the approach is insensitive to object color effects, as this example shows.

There was no similar analysis or discussion in the paper with regards to this.

This appears to be an implementation fault of using a Bayer pattern and not a fault of the main approach (using three BW phase shifted fringes, averaging them for color texture, and doing point-by-point phase estimation). This would not occur if, for example, a 3CCD camera were used, since for each point at least one color component would be sensed.

I mention this not because I favor a 3CCD, which I agree is more expensive than a Bayer camera; but thinking of the problem in terms of what actually makes the system work versus what is implementation brings up many questions about the applicability, limitations, and trade-offs of the system, as well as what is novel and the real contribution.

pages 14 and on have almost no changes from initial paper. Please see Rebuttal Comments for suggestions for these sections, especially advantages/disadvantages.

REBUTTAL COMMENTS

rebut pg 5. "We added the following statement in Section 1 at the end of page 6 ..."

It should probably go in the setup section, but I don't think the amount of detail here is that necessary. Just saying the computer, graphics card, and projector are adjusted to ensure approximate white balance is enough.

pg 5. "we added the following statement in Section 1 paragraph 6..."
You didn't mention using B/W fringes to avoid color bleeding. You didn't mention that averaging the 3 frames together removes the fringe pattern and allows you to get color texture information. This is also another key to your process. This should all be mentioned in the principles section.

pg 6. "3CCD camera could be used..."
My main point is that there needs to be a separation between what makes the technique work versus how you chose to implement it. The 3CCD camera just shows that there are other ways to implement the technique.

I think since many people will be interested in this paper, so it should be made as stand-alone and thorough as possible.

page 7. " Some slight problem could prevent the technology from advancing..."
That is why you need to highlight what step was novel and describe what is principle and what is implementation. To me, the use of the BW fringes, a point-sampling phase algorithm, and averaging to remove the fringes and get a color texture on a color sensor, solves the problem. It could be implemented in several ways - bayer, 3ccd, foveon, etc. The Bayer pattern makes it inexpensive and is compatible with the technique. It is cheaper than 3ccd and foveon, and doesn't have potential misalignment that the 3ccd has (however the alignment is done at the factory and is usually subpixel accurate). Explain that in the discussion section.

page 7. "We stated the B/W fringe, Bayer ... in our first version".
Yes, but you didn't explain in one place what allowed you to simultaneously solve the color bleeding/object color problems while simultaneously extracting registered 3D and texture. Nor did you discriminate between what actually solved the problem versus the implementation of that technique.

The explanation makes it clearer, but some of this should be in the setup section ("... The color calibration procedure is required ... we simply generate fringe images with R=G=B and adjust the color balance") and other parts in the principle section ("because the surface affects each B/W fringe ..")

page 8. "The calibration should be done in ... the setup section, because it isn't key to the operating principle. I don't agree with this ..."

The CALIBRATION details (not Bayer) should probably be in System Setup. In the principle section, you can just say that calibration is need to convert the phase to absolute coordinates; then in the system setup say you use the calibration approach with the red/blue checkerboard.

"The Bayer technology is very important... it should be addressed before the phase shifting ... since it is used when we discuss the technique in the phase-shifting section."

As mentioned in the review, the Bayer SHOULD go in the Principle section, but I still feel it should come after the phase shifting technology. Contrary to the rebuttal, the Bayer isn't needed to discuss the phase shifting technology. The phase shifting technology part of the principle section doesn't even mention the Bayer, just that intensities are used. You can mention color component intensities, then leave the exact implementation for the next part. I would think that the phase being computed from point intensities that are insensitive to object color and color bleeding should be stated, then say the Bayer pattern makes it inexpensive, without alignment issues (due to the filters being directly fabricated on the CCD), and is compatible with the technique.

"since the Bayer technology is important to the success of this proposed technique."

Using the Bayer does not solve the color bleeding problem, object color problem nor the fast simultaneous high resolution 3D/texture capture in one camera. The BW phase shifting fringes, point phase computation, and averaging to get color texture solves the problem. The Bayer is just an implementation of that technique to use a single chip (what about a single chip Fovean?) for spatially multiplexing color and making it cheap.

What is the most important part of your paper, one camera 3D plus texture without color effects OR using a single chip camera instead of a three chip camera? I believe it to be the former. In your rebuttal comments you too say "The key of this paper is to obtain precisely aligned 3D geometry and color texture".

page 8. "We believe that showing local noise problem in the discussion section is more proper since it is not a key point to the paper"

It's not a matter of whether its key to the paper or not. Its an experiment in the discussion section. The experiment section has both experiment and discussion, and the discussion has an experiment. Its more a matter of style, so I won't press the issue.

page 9. "Refer to Pros. above"
The only reason I suggested referring to Zhang, Tower, Tower papers is because they appear to be on a similar parallel track. So, it would be good to compare and show the advantages of your setup, of which you have many.

"I don't know which paper they address our technique ... I guess the reviewer might mistakenly regard somebody else's paper as ours."

Zhang, Tower, Tower "Shape and colour measurement" [2007], reference Peisen Huang's paper that uses color encoded fringes that is the center of several works done by Zhang S including this one. Zhang Z's [2008] paper also mentions Huang and collaborators. "Huang et al proposed a colour-encoded fringe projection technique using three patterns...". So, they do discuss the general track of Huang and collaborators, which Zhang S was one of, although not on the specific paper they referenced. I was saying that Zhang Z acknowledges your parallel track and compares/contrasts with it, so it might be useful to the readers for you to do likewise.

page 9. "If the projector ... only projects red spectra..."
That is an extreme example though. Color bleeding may actually help in this case, since your red projection may spill over into the green/blue spectra so you can still get an estimate.

A more realistic case is if the computer/card/projector is "off-white" or not carefully white-balanced, you'll still get a consistent phase measurement (no color bleeding or object color problems) although the texture may be not be white-balanced. Otherwise, as I pointed out in my review and you mentioned elsewhere in your rebuttal (but not in your paper), your system is not severely affected by white balance problems. This is the point I'm trying to make and think you should explain and emphasize. (I'm trying to help you out).

page 9. This is the reason I suggested referring to Zhang, Tower, Tower. This is what I see as a comparable work, so it's important to compare/contrast. If you don't think it's a problem, say so in the paper.

But if you have chromatic aberration (in which some cheaper projector lenses), the sine waves won't have the same frequency and the 3 phases shift images won't be consistent. If the 3 colored sinewaves are the same frequency but not aligned (the dual of the 3CCD camera misalignment case), then the 3 phase shift images also won't be consistent.

page 10. "we added the following statement in section 1 on page 6"
This is an important advantage that should be addressed in the either the Principle of the Discussion Section. This is why I suggested that you present the fringe pattern section first without the Bayer, then add the Bayer and show its advantages to other possible implementations.

page 10. "we added the price ... in Section 4 on page 12, experiments."
It might go better in Setup.

You still could've aligned the camera so the checkerboard wasn't rotated in the image. That's not a camera price issue. It's not important though.

You should talk about the robustness to lighting, barrel distortion, etc, that we see in the image. Use the image to your advantage. That is important.

page 11. "The noises are mostly caused by the sensors..."
Sounds like something to mention in the Discussion.

But is it "noise"? I see the checkerboard image and the mountains in the 3D renderings (without texture). It probably comes from where you have black in your texture, then you can't make a phase estimate. So what do you do? Just smooth it out with a gaussian? That would sacrifice the good parts. You could identify those areas (by looking for black in the texture) and smooth out the 3D using Poisson/gradient methods from the edges. Again, something to add to the Discussion.
page 11. "The measurement area of approximately ..."
This should be in the experiment section. At least in the caption of the image.

Actually, in my original review I was thinking in terms of measuring absolute position with respect to the camera, but this is still absolute measurements with respect to a target plane. I would like to know the distance to the camera though.

"You need to obtain something to sacrifice something else..."
These are the kind of tradeoffs people might be interested in. These are just suggestions of possible areas of discussion that you could chose from.

page 12. "Well the reader needs to read our previous papers."

I disagree that you should make a vague statement and expect the reader to either know your previous research or go dig out your previous papers to try to figure out what you are talking about. They can look up references to get a deeper understanding, but you have to give them reason to want to go look it up and guide them on how they are related. Without that, it makes the paper feel disjointed and adds to reader confusion.

Here's what caused the confusion.
In the introduction, you also state that realtime acquisition is one of the advantages of this approach. In the setup section you say that you acquire at 70fps. You mention in the experiments that the data acquisition is fast -- 1/4 second, but don't draw attention to or explain that you aren't using the full frame rate of the camera, so it can be easily missed. Then back in the Fast speed example, you say it can be used in your real-time system.

Its confusing because you don't say what the difference between this system and the realtime system is, and why or what changes might be made and how much improvement might be expected. The definition of real-time is application dependent but typically in motion cases either >15fps or >24fps is reasonable. I originally thought the paper example was "real-time", based upon earlier statements in the paper, especially that the camera running at 70fps. 70fps divided by 3 frames is still greater than 24 fps, but you don't make it emphasize that it gets reduced down to 4fps or why, so its no longer real-time.

Just a few expanded sentences to give the reader a background is enough.

All you need to do is say "Because only 3 fringe images are used for our measurement, this technique can be used for real-time 3D shape measurement (>15fps). In this paper an image set is acquired in 1/4 second or 4fps. If the B/W camera used in our previously developed realtime system[18] is replaced with a single-chip color camera synchronized to a DLP projector with the color wheel removed, simultaneous 3-D geometry and color texture acquisition can be reached in real time (up to 26.7fps or 120fps)."

The synchronization is important as well as the increase in frame rate. This should be mentioned.

page 12. "I think I have addressed this in the introduction ...
I still believe you should be comparing comparable systems (geometry + texture), not comparable costs.
I think it should be included in the discussion with your advantages/disadvantages.

page 13. "Well, these arguments are even weaker because the hardware costs are different.

I disagree that they are weaker. Again, you should be comparing comparable systems, not comparable costs.

Your arguments in this paper are for a system that gets 3D and texture simultaneously, and there are other systems including your real-time system that do that. The first point in the advantages was "it is very difficult to achieve
precise alignment if multiple cameras are used." So you are comparing multiple cameras. Even in your abstract you are comparing with 2 camera systems. You shouldn't compare it to systems that only find 3D, but to systems that do both 3D and texture. In this case you have an even bigger advantage!.

If you want to state it as 3D with texture for free, then state it in that manner; although I'm sure people will also say that its therefore even less than for two cameras. You can frame it the other way ... it less than for two cameras; its the same price as using a BW camera for 3D only but with texture for free.

If you don't state what you are comparing to, then I have to assume its the same as the previous comparison. Unfortunately, in the advantage/disadvantage section you are changing what you are comparing to and not saying that you are changing or why. It makes it very confusing or jolting.

page 13. "Are you only selling for Zhang ZH's work?"

No, I am not selling Zhang ZH's work. I do not know him nor his collaborators. I know of their work however, which is parallel with a comparable solution. Therefore, I must judge your work based on current published works in the field. I must also judge not just the ideas, but also the experiments, analysis, discussion and conclusions. Furthermore, I must compare the depth of your experiments and analysis with the standards of a peer reviewed journal.

Also if you look in the pros section, you can see that I prefer your technique (although I understand why they are going for their 4 fringe technique).

page 13. "light loss"
The sentence in the original manuscript's disadvantage section was short and didn't express the problem clearly or expand upon it.

Yes, in a 3CCD, 1/3 of the light also falls on each sensor, but there are 3 sensors, so very little light is wasted, rather than being absorbed like in the bayer filter. So, there are things you can do with that additional information, such as noise reduction if you are only interested in intensity, or other things I mentioned in the review.

You don't have to state this though.

I think its just wording. Start with the basis of the comparison and what you are comparing to, then tell its disadvantage.

"Light levels or exposure are important in a camera's ability to capture good fringe images. When comparing the Bayer camera to a B/W camera, the Bayer pixel has approximately 2/3 less light per pixel due to the absorption by the color filters. Therefore longer exposure time or brighter lights are needed for a Bayer camera."

Its a parallel argument, since you are restricting it only to viewing the fringes. And you are restricting it to a Bayer camera not all color cameras, so you are showing the weakness lies with the Bayer pattern not the technique (which you should also state)

page 14. "The whole point of this paper: low cost, simultaneous [3d plus color], and possibly real time..."

You forgot avoiding color bleeding and object color problems. These goals should be stated clearly in the introduction.

The low cost was not stated as a main selling point in the original paper in either the title, abstract or introduction. It was mentioned in the advantages, but in relation to a BW camera for fringe capture. The low cost could be also compared to a BW camera for fringe with a color camera for texture. You need to mention this in the introduction
(which you have in the revised version) and in the abstract (which you have not). It might even go in the title if its that important.

The suggestion of a 3CCD camera is merely to express that the main novelty is not the Bayer pattern. A Bayer pattern is merely low cost spatially multiplexed color sensing, and your choice of phase calculation is compatible with it.

Yes, a 3CCD may have problems (although the misalignment disadvantage is overstated), but it may also have other advantages. Regardless, the implementation is not the novelty.

In the original paper the novelty was not clearly stated and laid out. The concentration on the Bayer pattern made the confusion greater.

With the inclusion of the description of the approach, the advantages of the Bayer pattern become clearer, however, it is still not laid out for the reader. There should be a distinction between the advantages of your approach versus other approaches, and the advantages of the Bayer implementation versus other possible implementations.

page 14. "I am pretty sure that our paper provided an elegant solution ... even though we may not be able to convince you the paper is important"

I stated in my pros, I thought it was an elegant and simple solution and showed my admiration for it. That is beside the point. I'm trying to help you present your paper in the best light possible, so that people can understand it clearly, appreciate it, and gauge how well it would work for their applications.

But again, look at my pro comments, and also my conclusion, where I state that this is an important result (which is why I didn't suggest it be rejected and why I'm putting a lot of effort into this review). I state high admiration for the end result of your work. However, I find it lacking in analysis and discussion. I'm trying to help you present and round out your work to bring it to the standard of a journal paper, to show off your results and have it read, used and referenced. I'm trying to build up your paper, not tear it down.