

# JUDGING STEREO



(Cartoon courtesy of Lee Pratt)

**A Presentation by**

**George Themelis**

**(w/contributions by Boris Starosta and Al Sieg)**

**Detroit (DSS) / September 2011**

# Outline

- Stereo Competitions—Why and How?
- Mechanics of Judging
- Essence of Judging
- What do the judges look for?
- How do better in stereo competitions
- Critique
- Assignments
- A Judging Experiment
- Questions, Comments, Open discussion

# Competitions—Why?

Why bother with competitions, judging, scores & awards?

- Process of **sharing pictures** involves some degree of judgment. Competitions is a step further where the process of sharing and judging is done in a formal way.
- One of the objectives of stereo organizations is to **encourage members to take pictures**. The purpose of competitions is to motivate members to take and share pictures and improve their photographic skills through constructive criticism.

# Competitions—How?

- Local level: Stereo Club (personal interaction)
- National/International Level: [PSA Stereo Exhibitions](#)

# Judging—Mechanics

- How Many Judges?
- Judges' qualifications?
- Best Scoring scale/range?
- One overall score or score in different areas?
- What makes a good judge?
- Discussion among judges?
- Should a judge's opinion agree with "majority opinion"?

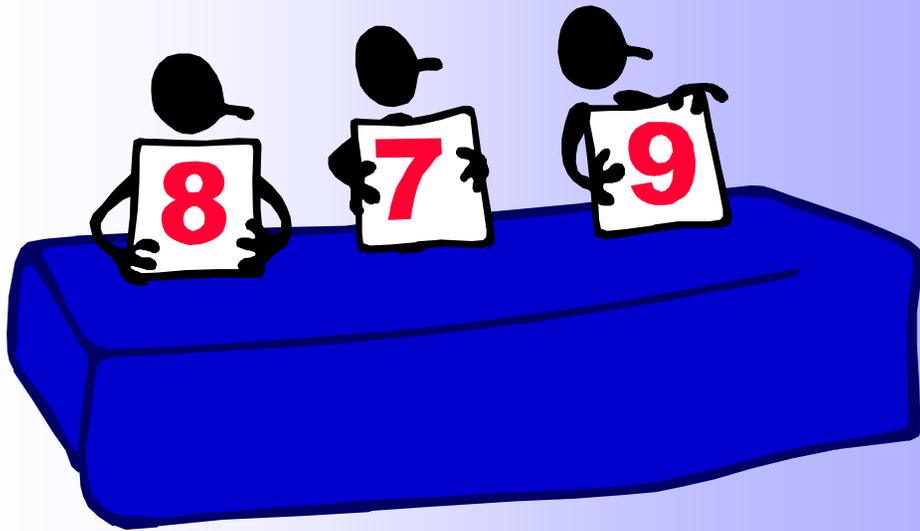
# Judging—Mechanics

**Competent and consistent judging is important for the success of any photographic competition**

*Inconsistent judging usually turns people off competitions*

- **Judge Qualification: Do judges need to have experience taking pictures on their own?**

Even though it is not a requirement, I think it helps if the judges are also stereo photographers. This helps them understand what is involved in making certain images, identify a creative approach, recognize the difficulties involved and how good the execution is, instead of just responding to the appeal (or lack of) of the images. Being a judge is a good experience for the judges too. It makes them better photographers and better judges of their own work.



- **Number of judges**

Some clubs use only one judge. Most clubs (our club too) use **three judges**. This leads to smoother results because individual preferences tend to average out and the average score better approximates a “popular vote”. There might be clubs that use more than three judges but this usually complicates the judging process.

Sometimes it is difficult to find judges who do not have entries in the competition. A solution is to have a 4<sup>th</sup> judge and then count this score in the place of the judge whose entry is being judged. This system works better if the 4<sup>th</sup> judge is asked to judge all entries without knowing when his vote counts or not. These systems work OK if all judges follow a similar scoring distribution.

## • Score Range

In the simplest form judging can be done using only two scores, say 0 and 1. 0 means “out” or “not accepted”. 1 means “in” or “accepted”. This system is usually used to select/accept or eliminate certain entries from a group of images. The next step is to add one more score, say 2, for “accepted with awards”. This system is suitable for an exhibition where each entry is either not accepted, or accepted, or accepted with awards.

I think that most judges cannot handle more than 3-4 levels, and too many choices lead to confusion. Many clubs use a **5 point scoring system**, either from 1 to 5, or from 5 to 9.

DSS this years is switching to a **6 point scoring system**. This is interpreted as follows:

<b>Score</b>	<b>Given for...</b>	
<b>9</b>	<b>Exceptional</b>	Flawless in all aspects; cannot be improved
<b>8</b>	<b>Outstanding</b>	Near-perfect, only minor improvement is possible
<b>7</b>	<b>Great</b>	Well made, award worthy
<b>6</b>	<b>Good</b>	Above average, still has room for improvement
<b>5</b>	<b>Average</b>	Common, typical “snapshot”
<b>4</b>	<b>Deficient</b>	Has significant flaws

- **Score Range (continued)**

No matter what the range is, **it is important to use the entire scoring range available**. A judge who does not use the entire range is making himself invisible, i.e. his/her score is lost and does not count. As an extreme, consider a judge who gives every image a 7 in a 5-9 range. This judge is as if he never judged! He has absolutely no effect on the results.

Here is another example: Consider a competition with a 1 to 9 range. Two judges decide to use only the upper part of this range (5-9) while the third judge is using the entire range. Image A is being judged. This is liked by the two judges who give it 8 (worthy of award). The third judge does not like it and he gives a 3. Total score for image A: 19. Then image B is being judged. The situation is reversed. The first two judges give it a 6 (low score) while the third judge gives an 8. Total score for image B: 20. So image B got a higher score than image A, even though the majority (2/3) of the judges liked A more than B. **The third judge's opinion weighs more because he is using the entire scoring range.**

If you find yourself judging an exhibition with 1 to 9 scoring range while you want to use a 5 point range, instead of using the upper end of the scale (5-9), it is better to spread your 5 points over the entire range, for example: 1, 3, 5, 7, 9.

- **Consistency in Judging**

The most important characteristic of a judge, in my opinion, is **consistency**. It does not matter if his/her scores tend to be high or low, as long as they are consistent, i.e. do not change with time and the same quality of image receives the same score, no matter where it is placed in the judging sequence.

This is especially important in the first few images when we naturally tend to score them as “average”, regardless of their merit, thinking that will allow room for both higher and lower scores later.

- **Preview or not?**

Experienced judges have set a standard in their minds through many judging sessions and use it this standard as a reference, not the previous images. For inexperienced judges it is a good idea to first **preview the images**. This will allow them to make better use of the entire scoring range. If preview of all images is not possible because of time constraints, then previewing just the first few slides can help for the judges to establish a scoring scale.

- **Discussion among judges?**

Should Judges discuss the top scoring images before giving a top award (Gold Medal, Image of the Year, etc) or accept the results of the scoring (highest scoring image gets the top award, no questions asked)?

**My strong opinion: YES**

- **Scoring Machines and other Judging details**

When images are judged in projection:

- Should titles be read?
- Should score be announced immediately?



# Judging—Essence

- **Technical vs. artistic qualities**
- **Use of depth?**
- **What do judges look for?**
- **Safe vs. controversial images**
- **How to improve your chances of doing well**

# Judging—Essence

Newcomers in stereo imaging with some experience with judging “flat” images are usually puzzled. They realize that there are two qualities in our stereo images:

1. **Photographic qualities that apply equally well to flat images**
2. **Use of stereo depth**

So they ask: *“How should I judge an image that shows good photographic qualities but does not have much depth, vs. an image that shows great use of depth but is not a great photographic image?”*

Similar questions are asked with regards to **technical vs. artistic qualities**.

*“How do I judge an image that shows great artistic qualities but it is mismounted?”*  
Or, *“How do I judge an image that has strong impact vs. one that does not have much impact but it is very creative?”*

Finally, if there is an **assignment** in the competition, I often hear this question:

*“How should I judge an image that barely meets the assignment vs. another image where the assignment is the focus of the picture?”*

Clearly, there are a number of aspects on a stereo image that need to be evaluated. I have made an attempt to classify these in the Table in the next page.

<b><u>Aspects of a Stereo Image to be taken into account during judging</u></b>	
<b><i>Technical</i></b>	<ul style="list-style-type: none"> <li>• Is the exposure and focus OK?</li> <li>• Is the stereo image properly mounted? (no distracting mounting errors and good use of the stereo window)</li> <li>• Is depth within acceptable range?</li> </ul>
<b><i>Artistic</i></b>	<ul style="list-style-type: none"> <li>• Is thoughtful composition shown?</li> <li>• Is a creative approach evident?</li> <li>• Does it show good use of depth?</li> </ul>
<b><i>Assignment</i></b>	<ul style="list-style-type: none"> <li>• Does the image meet the assignment?</li> <li>• Is the assignment a central point or just an afterthought?</li> </ul>

Certain clubs have an itemized list of qualities and each image is judged in all different qualities and then the scores are summed up for the final score of his image. In our stereo competitions we only give one score and judges are expected to take into account all these factors and come up with one score, within a reasonable but short amount of time. How is this possible?

I believe it is possible to look at an image and without much mental calculation or hesitation, assign a numerical score, summarizing all these values. This is possible if there are not many different scoring choices, as in the 5 point range scale. This is the result of some experience in seeing and judging stereo images.

Of course, this numerical score is only a **subjective evaluation**. By nature, judging cannot be objective or we would have computer software doing the judging. Different judges interpret different images in different ways. Those who have watched the scores from three different judges usually comment that many times all three judges agree or are within one point range from each other, but there are cases of wide variations in some images. That's part of human nature. No one is right or wrong. It is a matter of subjective interpretation.

# Personal Thoughts on Judging

Going back to our first question of **depth vs. photographic qualities**, this is something that each judge has to answer on his/her own. I cannot offer any set rules for weighing these two different qualities. Certain people (usually beginners) are very impressed by the use of strong depth in stereo pictures. There are people who believe that use of depth should be the most important consideration in a stereo image. After years of dealing with stereo images I tend to put depth in equal footing with other photographic qualities. I have no hesitation to give a high score to a good photographic image that shows only moderate to low depth.

How about mounting and stereo window placement, areas where a beginner might have a problem? Again, I have no set rules but I do take these elements into consideration. If the mounting error is distracting (hurting the eyes) I will take points off. If the stereo window could have been improved (a subjective call) I might take a point off, but it depends on the image. I will usually reward an image with a good placement of the stereo window with an extra point, rather than take a point off for questionable window placement, unless if this placement is a clear and distracting “stereo window violation”.

If the competition has an assignment I usually reward an image where the assignment is a clear and well-placed element of the picture, rather than take points off when the assignment appears to be a coincidence or afterthought.

I generally reward images if I think they are creative, required a lot of planning and the execution is good vs. images of the type “been there, taken this shot”. Humor, creative approach, good execution, all these factors count.

# Personal Thoughts (cont)

I remember one picture I entered in Detroit when the subject was “Music”. I took a picture of my 4 year old niece playing the piano. I went through a lot of trouble to take this picture. First, it took a lot of planning. I had a clear image of what I was trying to achieve and how to go about it. Despite my wife's objections, I climbed up a step ladder to get a higher perspective. I used all kinds of tricks to make her smile. After I took the picture, I spent considerable time mounting. I used double masking to eliminate distractions and make the keyboard “go through the window”. I rarely work that hard for a stereo picture! Result: The picture scored poorly in Detroit. The judges called it a snapshot. They pointed to the harsh on-camera flash and some distraction (a plug in the wall) in the background.

The moral of this story: It is not enough to work hard to create an image. You also have to convey this through the picture. The judges do not know how was this picture taken. They judge from what they see and clearly this picture did not convey the work that went behind it.

Occasionally, a **new technique** or a group of images appears in stereo competitions and these images tend to score high in virtue of their uniqueness. This I think happened at some point with hyperstereos. And more recently with HDR stereo pictures. As these images become more and more common, the standards get higher, meaning that the subject alone is not enough for the image to score high. It has to show good execution and other qualities that are expected from more traditional images. This is, I believe, a good thing. On the other hand, more traditional images (scenics for example) are so common that it takes an extra special image to do well.

Judging a wide variety of photographic topics (scenics, pictures of people, macros, etc) with the same set of standards is a bit difficult at first but exposure to a variety of stereo images, both as a judge and as a participant in competitions, helps to put things into proper perspective.

# What do Judges Look For?

By Boris Starosta

- Does the work advance the state of the art: is there something new in it?
- Does it explore unusual techniques or applications in stereoscopy? (time exposures, rivalry effects, spatial/scale control, non-synchronicity, etc.)
- Does it explore effects/subjects uniquely suited to stereoscopy? (depth, iridescence, complexity, camouflage, rivalry, transparency, etc.)
- Does it have stereo-aesthetic merit (e.g. use of stereo window)?
- Other aesthetic merit (color, positive/negative space, textures, rythm, composition).
- Does the image move the emotions?
- Does the image make a political statement?
- Is the image dated, or timeless?
- Is there humor in the image?
- Good craftsmanship? (focus, stereobase, exposure, alignment, stereo window, etc.)

# How to Improve your Chances

By Al Sieg

- Pay attention to **technical details** (exposure, mounting, etc)
- If possible **preview your photos using the same viewing method used for the selection of the photos** in the exhibition (for example, if entries are projected, view your photos in projection)
- **Attend as many exhibitions as possible** to get a feeling of what is accepted and see a variety of “safe” and “controversial” pictures.
- Key element in a photo: **Light**. Best scenic pictures are taken early or late in the day.
- **Choice of subject**: Your pictures are competing against other possibly similar pictures and the judges have already seen common subjects many times. Your chances of acceptance improve with an unusual subject.
- **Fads come and go**. A new technique (hyperstereo, blacklight etc) at first gets “easy” acceptances but as it becomes more common, it becomes more difficult to do well.
- Bottom Line: You must make a photograph that stands out of the pack in order to increase your chances of acceptance in an exhibition.

# Critique is Important

Judging usually assigns a numerical score to an image. Critique is the process that analyzes the strong or weak points of this image and attempts to explain the score attached to it. Critique is important because it goes into details and gives good feedback on how to improve pictures. Some forms of exhibitions/competitions/image sharing do not have any scoring, just critique in the form of comments.

Certain people through education and experience are very good in critiquing images. I am not one of these people. I tend to react to an image by indicating whether I like it or not, but I am not good in explaining with words my reaction to certain pictures and I cannot verbalize well what is it that I like or don't like. When I hear fellow judges comment I usually agree and think "that's exactly how I feel too".

In any case, it is nice to have people in a club (not necessarily serving as judges) who can offer a meaningful, thoughtful and constructive critique for at least some of the stereo pictures we see in the competitions. This makes the process valuable not only for the person whose image is being critiqued but also for the rest of us who are learning from these comments.

# Assignments

Not all stereo clubs have assignments as Detroit does. I think this is one of the best features of our stereo slide competitions.

Photographers tend to develop favorite subjects. Some photographers for example like to take pictures of flowers. Without a system of assignments for motivation, one could continue shooting only his or her favorite subjects. I see the assignments as a challenge to explore a new area, a new topic, a new subject.

I understand that some people with a lot of years of stereo photography, instead of trying to take new pictures to cover a given assignment, they just go through their inventory of stereo slides and pick a few that match the assignment. I am guilty of doing the same thing. This defeats the purpose and reduces the effectiveness of the system. Next time, try to take new pictures for a given assignment.

# Judging Experiment

## The Experiment

During our October 2003 meeting we conducted a judging experiment. First, we explained the mechanics and essence of judging and the 5 to 9 scoring scale. Then 24 slides were projected. Everyone in attendance was given a “report card” and asked to score each projected slide using the 5 to 9 scale. 35 people participated.

The slides came from various people, none of whom was present at the meeting or known to our club members. An effort was made to include a “random” variety of images, some good, some average, some not so good.

## First Reaction

During my correspondence with a friend, not a club member, I mentioned this experiment and he said: *“I bet you will be disappointed with the results”*. This comment sounded odd to me. Why would I be disappointed? This is an experiment, I have no expectations, and I am looking forward to analyzing the results.

But when I first looked at the results, I was disappointed! I marked the high and low score given to each slide and I was amazed to see that 16 slides were given at least one 5 and one 9! In other words, for  $\frac{3}{4}$  of the slides someone in the room thought that this was one of the best and someone else thought this was one of the worse slides they had seen!!! What does that say about agreement among different judges?

Looking closer at the score distributions I realized that the situation is not too bad because there are a lot of judges and usually there is just one person that went to the extreme and the majority agreed within a closer range. But still, I was expecting a better agreement. I suppose I am a bit disappointed because I believed that there is a better agreement between different people regarding what is a good and a not so good picture.



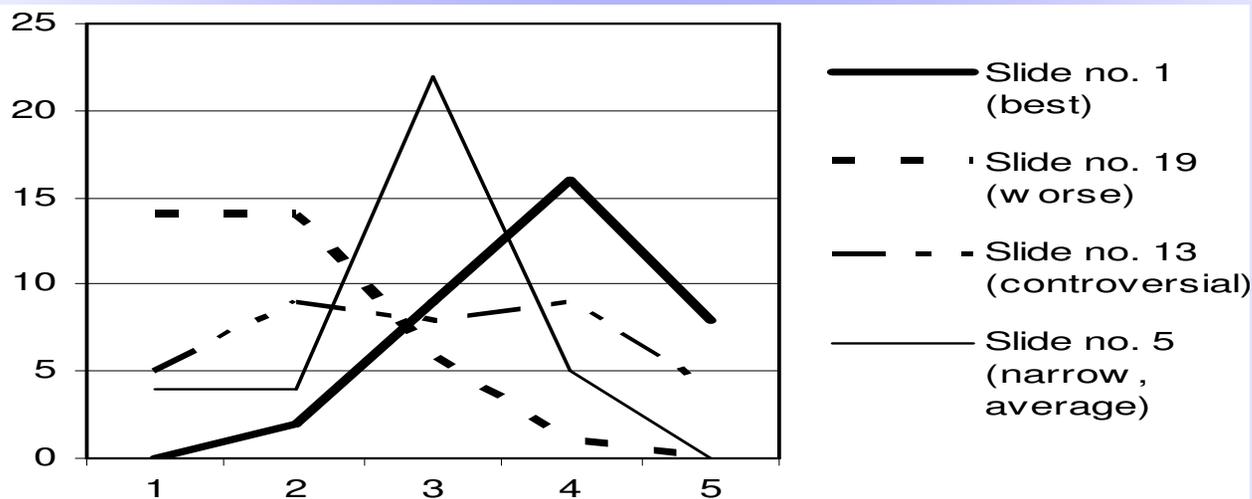
# Judging Experiment (cont)

While it is common for people to disagree about the very best pictures, there is usually an agreement about the very worst pictures. I tried to find one of the worse stereo slides imaginable to include on purpose. This is a scenic that is overexposed, out of focus, with bad composition, window violation, washed colors, not much depth, and, to add insult to the injury, it was scratched in several places. I thought to myself, “*if this does not get a 5, then nothing should get a 5!*”

Looking at the results we see that this is the lowest scoring slide, as expected. However one person gave it an 8 and 6 people gave it a 7. These same people gave lower scores to other slides. That's really puzzling for me. I had really expected to only see 5s and 6s for this one.

## Controversial Pictures

Most people usually look only at the average (or total score) but in this experiment we can also look at the standard deviation for each picture. A low standard deviation means that there was a good agreement among different judges, while a high standard deviation means that there was a wider variety of opinions, i.e. the picture is controversial.

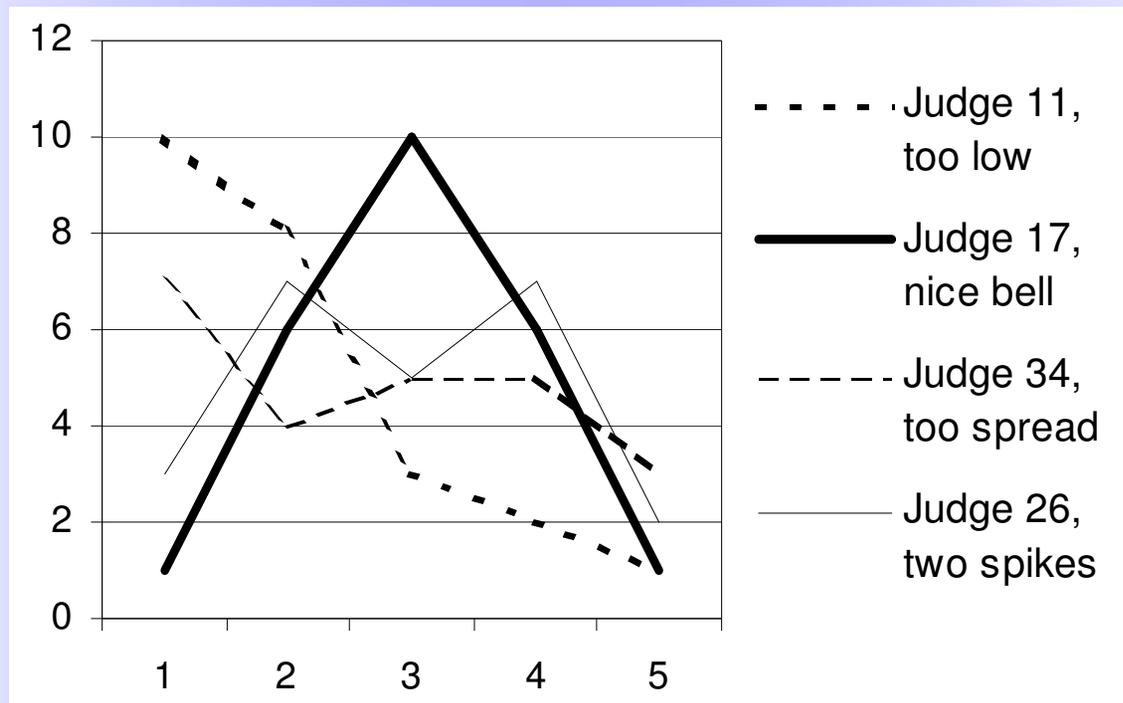


# Judging Experiment (cont)

## Judging the Judges

Judges can also be characterized by their average score and their standard deviation. The average score should be around 7. Higher or lower averages show a tendency to score higher or lower overall. Extremes in average values should be avoided (assuming that the pictures judged are randomly distributed and not selected to be all very good or all very bad on purpose.)

The standard deviation shows how wide a judge is spreading his/her scores. Both low or high STD values indicate a problem. If a judge gives every slide the same score, say 7, then the standard deviation is zero. If a judge gives random scores (i.e. every score has equal probability) then the standard deviation is around 1.4. A value around 1.0 indicates a “healthy” spread.



# Judging Experiment (cont)

## The Majority Opinion

A judge's score distribution does not tell the entire story however. A number is needed to indicate how well a judge agrees with the “majority opinion”. The majority opinion is the average score given to each slide by ALL judges. To characterize this agreement (or lack of) with the majority I created my own number which I call “deviation from the majority” (DEV). Small DEV values indicate that a judge agrees with the majority opinion, while large DEV values indicate that the judge's judgment deviates from the majority opinion.

## What is a “Good Judge”?

We have already mentioned some characteristics a good judge should have, like consistency. A “healthy” scoring distribution (not too low or too high scores, distributed on a bell-shaped curve, not too narrow or too spread out) is another desirable characteristic, assuming always that what is judged is a group of slides whose quality is also similarly distributed. Thanks to what is known as the “law of averages” the more pictures are being judged, the closer the scoring distribution of a given judge should be to this nice bell-shaped curve.

In addition to the shape of the scoring distribution for each judge, this study also compares individual judgment vs. “majority opinion”. This raises the question: **Should agreement with the majority be a desirable characteristic for a judge?**

I hesitate to say that it should be, considering that judgment is something personal and subjective. But, on the other hand, would you want a judge whose scoring is no different than pulling random numbers out of a hat? While it is expected that in certain instances a judge might disagree with the majority, it is not a good sign, in my opinion, if this disagreement is consistent. Inconsistencies in judging and wide deviations from the majority opinion tend to create friction among the participants. I think that experience both in judging and working on the medium judged, reduces these inconsistencies and makes judging more fun and less frustrating.

# Summary

- The purpose of stereo competitions is to motivate people to take and share stereo pictures and improve their photographic skills
- Assignments, judging, and awards aim to challenge, motivate and reward members for their photographic efforts, and make the sharing process more interesting for those who participate and those who do not.
- By nature, judging is subjective. There are a lot of aspects that need to be evaluated within a short time. Through experience each judge has a mental image of what each score stands for and he/she should be able to assign a score to each image.
- Consistency is the most important quality of judging. A judges' scores should not go up or down with time. They should stay consistent and an image should receive the same score, regardless of its order in the judging sequence. A judge should try to score in the entire range and not be afraid of the extremes.
- Judging stereo images is a great experience that helps a person better evaluate his/her own images, not just those of others.
- Constructive critique of images is an important supplement of judging that helps photographers improve their skills. Some people are better in critiquing than others.
- To increase your chances for doing well in competitions you must make photographs that stand out of the pack in choice or subject matter or approach to photographing a given subject.