A Nonparametric Statistical Approach for Stereo Correspondence

Sema Candemir and Yusuf Sinan Akgul

***Abstract*— This paper introduces a novel non-parametric statistical metric that can decide if the recovered set of parameters from a Computer Vision optimization process can actually be considered as a statistically significant solution. The level of significance can be used as a quality metric of the solution which makes it possible (i) to compare the solutions obtained using different optimization methods, and also (ii) to set intuitive thresholds on the acceptance criteria. We chose the stereo correspondence optimization methods as the initial test bed for the new technique. We compare and combine the results of Sum of Squared Differences (SSD) and Sum of Absolute Differences (SAD) methods for stereo correspondence. We validated our claims by running experiments on standard stereo pairs with ground truth. The results showed that the introduced ideas actually work very well and they can be used to improve the optimization results from different sources.**

I. Introduction

The most general task of Computer Vision is to extract parameters of the real world objects from images or image sequences using a model which is usually geometric, physical, or statistical. The parameter extraction process can also be viewed as imposing model related restrictions on the images and retrieving the best set of parameters that conform to these restrictions. The degree of efficiency and applicability of a Computer Vision model is comparable to the convenience of imposing model related constraints on the images and how strict the restrictions can be made. If there is a well defined method of imposing the model constraints on the images, then the model parameters can be easily obtained. Similarly, if the constraints can be made very strict on the images, then the model parameters can be recovered even with very noisy images.

In Computer Vision, the efficient and applicable recovery of the model parameters are generally done by function optimizations. Correctly optimized functions produce the best set of parameters for the system model that define the solution on the current images or image sequences. However, the best set of parameters does not guarantee the existence of a correct solution to the problem. The optimization process cannot decide if there is no feasible solution for the given problem on the current images. Using a predetermined or dynamic threshold value on the optimized function values usually does not produce reliable results because the objective functions are very difficult to design to allow such stable thresholds. Furthermore, the optimized values are almost always dependent on the system parameters, i.e., if the objective function is optimized using a different parameter set, the resulting optimal function values change dramatically. As a result, using the objective function values to compare the quality of different solutions obtained through different parameters sets is not possible. The above problems about classical Computer Vision optimization make it very difficult to design generic systems: (i) one always needs to choose threshold parameters experimentally for each application, (ii) it is not possible to compare solutions from different optimization functions on the same problem, (iii) as a result it is not trivial to automatically combine solutions of different optimization processes.

In this paper, we propose a non-parametric statistical method that measures the statistical significance of the results found by the Computer Vision optimization methods. Since the quality of the results are measured using the same technique regardless of the optimized function types, it is possible to compare and combine the results from different optimization processes. In addition, the proposed solution quality metric can be conveniently used to set thresholds on the optimization results in a meaningful and intuitive way that would eliminate any error-prone try-and-observe threshold selection procedures.

Comparing and combining different optimization functions are employed before by a few Computer Vision researchers. The most common one is the weighted sum method, which is used popularly in many regularization based methods such as snakes and graph cuts where data and smoothness terms are combined using weighs. This procedure introduces a new weight parameter to the system and it does not guarantee stability. Researchers proposed different methods to estimate the weight parameters. In perceptual grouping [1], Sarkar and Soundararajan [2] proposed a learning algorithm to decide the relative importance of the salient relationships such as proximity and continuity. In the stereo system of Klaus *et al.* [3], a self-adapting dissimilarity measure is employed to combine the Sum of Absolute Differences (SAD) method with a gradient based measure to find the correspondences. This heuristic method finds the desired weight by maximizing the number of reliable correspondences between the views. Although similarities exists between our method and the other methods, there are fundamental differences. First, our method is not based on heuristics. Second, our method is easy to implement and understand because it is based on standard statistical hypothesis testing mechanisms. Most importantly, our method can be extended for any application regardless of the Computer Vision problem domain, which makes our system very flexible and generic.

S. Candemir is with GIT Vision Lab, Department Of Computer Engineering, Gebze Institute Of Technology, Gebze, Kocaeli 41400, TURKEY `scandemir@bilmuh.gyte.edu.tr`

Y. S. Akgul is with GIT Vision Lab, Department Of Computer Engineering, Gebze Institute Of Technology, Gebze, Kocaeli 41400, TURKEY `akgul@bilmuh.gyte.edu.tr`

Although the introduced fundamental idea is conveniently applicable to other Computer Vision areas, we choose to apply our technique to the classical stereo correspondence as the initial test bed environment. We choose two popular stereo correspondence optimization methods: Sum of Squared Differences (SSD) and Sum of Absolute Differences (SAD). We analyze their results in terms of their statistical significance using our new quality metric technique. This analysis makes it possible to judge whether the optimized solution can be considered as a valid solution. In addition, for each pixel correspondence, it is also possible to choose the statistically more meaningful solution among the two solutions.

The paper is organized as follows. Section II introduces the background about non-parametric tests and their applications in Computer Vision. Section III introduces the fundamental ideas of our technique and its application to stereo correspondence. We describe the experiments and the validation of our work in section IV. Finally, we provide concluding remarks and future directions in Section V.

II. Non-parametric Tests

Our task of measuring the quality of an optimization process requires generality as well as robustness. The task of comparing and combining the results of different optimization processes requires even more generality and robustness. The standard statistical hypothesis testing can be used to compare results in a robust way with a given statistical model at hand. However, due to the generality requirement, we cannot assume a statistical model for our case, which forces us to choose a non-parametric statistical method to compare optimization results. Specifically, we employ permutation tests which can provide us with the required generality as well as the robustness features.

Statistical significance is a probability value, which is a measurement whether the outcome can happen accidentally or not. The permutation test is a type of statistical significance test and it was first proposed by Fisher [4] and Pitman [5]. Distribution of the observation are generated by using all permutations of data rather than from formulas and so it can be used in applications where traditional tests do not apply. The permutation test generates a reference distribution by means of randomization. This reference distribution is then used to asses the significance of observed statistic. Statistical significance of observed ordination is calculated based on the location of the statistic on distribution of the ordination. A value in the tail would rarely occur by chance and so there is evidence that something other than chance is operating. The generic way of forming the distributions of the permutation tests makes them very suitable for our purposes because we do not like to make any assumptions about the optimization process for which we would like to produce a statistical significance value.

While computing the distribution of the statistic of the permutation test, using all permutations of the data will take exponential time, especially when the data size is large. This makes the method unpractical for usage in the real time applications. Instead of using all permutations, shuffling the data and creating different samples from the complete permutation set can be used. Each shuffle generates one permutation of the data. Empirical distribution of the test statistic have been generated by calculating statistic for each shuffle. This procedure is called Monte Carlo test or approximate randomization test [6].

When the permutation tests first appeared in the 1930's, the high cost of calculating all the permutations made it unpractical in real world, although it was considered a powerful technique since it did not need any assumption about the population. Later, approximate randomization techniques were adopted for distribution generation [6] instead of exhaustive permutations. Pagano [7] suggested polynomial time algorithms for computing the distributions. Although some researchers still work on generating distributions in an efficient way, the current state of the art computers made permutation tests increasingly feasible, even for large data sets.

It comes as a surprise to us that in Computer Vision, only a few researchers employed permutation tests for their work. Golland and Fisch [8] have used permutation test to measure statistical significance of classification accuracy. They estimated statistical significance of the observed classification accuracy by chance due to spurious correlation of the high-dimensional data patterns with the class labels in the given training set. Other medical image analysis studies which employ permutation test are Bullmore [9] and Nichols [10]. Bullmore used a permutation test to estimate statistical significance of individual voxels or clusters of voxels. Nichols has used permutation test for functional neuroimaging.

Although our new optimization result quality metric uses permutations tests like the above systems, our application of permutation tests is completely novel. We mainly use the permutation test to bring the results of different optimization processes into the same "comparable" base. As a result, along with taking all the advantages of permutation tests, we can conveniently compare and combine different optimization techniques in a very efficient and elegant way.

III. A new quality metric and application to stereo

Although establishing the stereo correspondence is one of the oldest problem of Computer Vision, it is still heavily studied. The recent introduction of a ground truth image base [11] made the research even more popular. The input to the problem of stereo correspondence is more than one image of the same scene V taken from different viewing angles (Figure 1). The stereo correspondence is defined as finding the corresponding points in the left and right images. For example, in Figure 1, point p_1 in the left image corresponds to point p_2 in the right image because they are both produced by the same real world point P in volume V. Once the correspondences are found, the position of the real world point P can be found using simple triangulation.

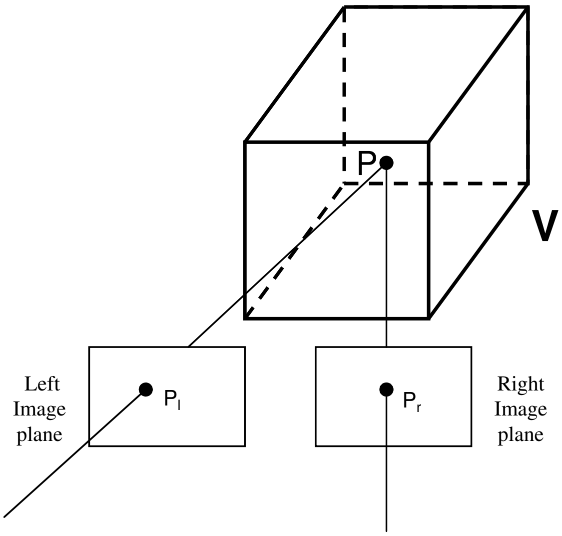

Fig. 1. Stereo image pair of the scene volume V.

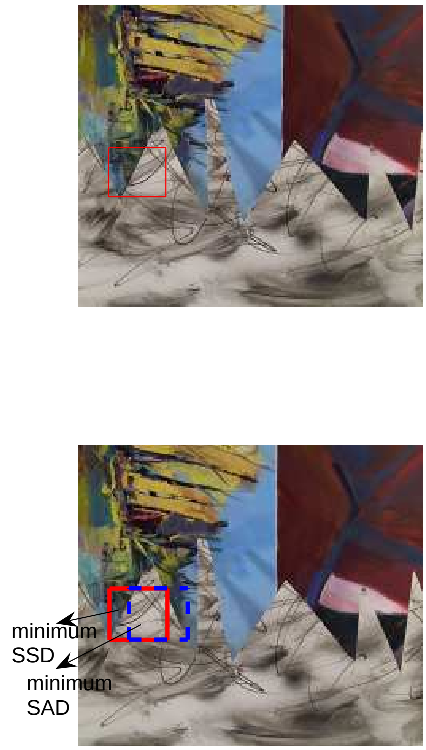

Fig. 2. Sawtooth image pair from Middlebury image base [11]. Up:Left camera image. Down: right camera image. The blue and red boxes in the lower image corresponds to the red box in the upper image if we use SSD and SAD, respectively. The optimal SSD value is 956 and the optimal SAD value is 135.

The offset between the image positions of the corresponding points in the two images is called disparity, and it is inversely proportional with the depth, which is an important knowledge of 3D scene. The general dense stereo algorithms match individual pixels in corresponding scanline pairs by optimizing image similarity metrics. The literature includes many different objective functions such as SSD, SAD, cross correlation, and normalized cross correlation. The SSD method picks two candidate image regions for correspondence (such as the red boxes in Figure 2), calculates the square of the difference between each pixel and sums them up. The SSD result is small if the selected boxes are similar. The SAD method works the same way, except it takes the absolute values of the differences instead of squares. The stereo correspondence problem can be partially solved using SSD or SAD by choosing a match for each pixel that produces the smallest SSD or SAD value. For some images SSD works better and for others SAD is better. Since the optimal values of SSD and SAD are very different for the same pixel correspondence, we cannot know which one is more trustable without knowing the ground truth. Figure 2 shows the matches obtained using SSD and SAD and it cannot be decided which match is better by looking at the optimal SSD and SAD values. If the quality of the results are measured using the same general technique regardless of the optimized function types, comparing and combining the results from different optimization process would be possible.

The classical permutation test requires all possible problem inputs to be formed and solved. The solution values are used to form a new statistical distribution. Any new problem is solved and the result is looked up in this distribution for the significance. Forming the all possible inputs or subsets of inputs might be time consuming. In our case, we follow a fundamentally different path to form our permutation distribution, which results in greater efficiency and more robustness. For each pixel in the left image, we evaluate all possible matches in the right image and measure the similarity using SSD and SAD. The results of each similarity evaluation is a permutation resample which forms empirical distribution of the objective function on the observed image. We continue this process until all optimal SSD and SAD solutions are found. When this step is complete, our reference distribution is also calculated with a very small computational load, which is very similar to histogram formation. Figure 3 shows the obtained distributions of SSD and SAD optimization methods on the same random image pair. An observed SSD or SAD statistic (score) in the left tails of the distribution is a strong evidence that this statistic can rarely occur by chance (random variance). Similarly, a value in the main body of the distribution could easily occur by chance. A p-value indicating the statistical significance of an observation can easily be produced by locating the observation on the distribution and comparing the number of observations from the left side to the right side. For example, in Figure 2, the p-value of minimum SSD is calculated as 0.008, and the p-value of minimum SAD is calculated as 0.015. Since a smaller p-value indicates more significance (less random variability or chance), we choose the match obtained by SSD in this case. As can be seen very easily, this approach makes it very convenient to compare solutions from different optimization methods. It is also trivial to combine two different solutions by choosing the statistically more meaningful solution.

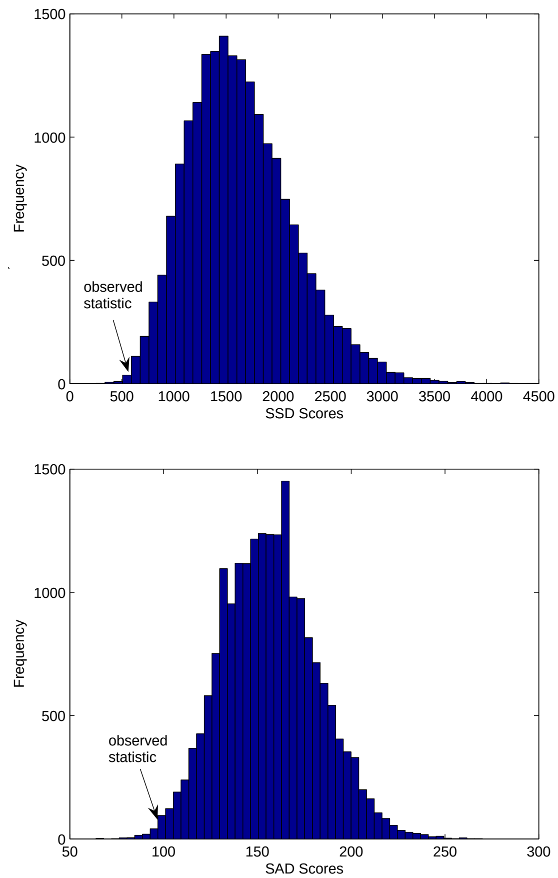

Fig. 3. Reference distribution of the SSD(top) and SAD(bottom) optimization functions. Location of the observed statistics on distributions produces p-values. We can compare and combine solutions by using these p-values.

It should be noted that our approach requires minimal extra computational load because in order to find the minimal SSD and SAD solutions, all possible cases have to be tested for the classic SSD or SAD based correspondence. The only additional computational load is to form distributions and deciding which solution is statistically more significant by looking up the distribution, which are both linear time operations.

Note also that the position of a statistical observation on the distribution can be used as a very intuitive threshold value. The threshold selection is a very popular technique in many Computer Vision systems. However, choosing a good threshold can only be done by experimentation in most of the systems because the function values do not necessarily make any sense for humans. For example, the SSD value of 956 in Figure 2 does not make any sense in terms of acceptability. However, the corresponding p-value of 0.008 indicates that such a match can occur by chance only 8 out of 1000 times, which is much more meaningful.

IV. Experiments

In order to validate our claims, we performed various experiments using the stereo test data with the ground truth from the Middlebury stereo base [11]. We run both SSD and SAD correspondence algorithms on different stereo image pairs and calculated the disparity maps for each algorithm. We also calculated the disparity maps for our method that combines SSD with SAD (SSD+SAD) using the new non-parametric statistical approach. We also changed the sizes of the match regions to eliminate any dependency on the window size parameter. Figure 4 shows the visual results from SSD, SAD, and SSD+SAD systems along with the ground truth for a sample image.

TABLE I

Number of incorrect matches for SSD, SAD, and our SSD+SAD methods

Image	Window Size	# of incorrect matches by SAD	# of incorrect matches by SSD	# of incorrect matches by SAD+SSD
barn1	7	32682	32056	31628
barn1	9	32296	32450	31538
barn1	11	32744	33334	32176
barn2	7	45682	43874	43968
barn2	9	43518	42408	42288
barn2	11	43067	42165	42092
bull	7	44596	41948	42101
bull	9	42217	40403	40120
bull	11	41553	40300	39830
cones	9	105961	104038	102805
cones	11	105009	104234	102666
cones	13	104292	105293	103643
map	5	19259	19437	19169
map	7	17814	17727	17538
map	9	17720	17371	17375
poster	9	47696	45885	45798
poster	11	46263	44865	44506
poster	13	45635	44786	43967
sawtooth	7	42609	42920	42228
sawtooth	9	42741	43751	42596
sawtooth	11	43866	45209	43863
teddy	7	87204	85502	84749
teddy	9	86839	85802	84332
teddy	11	87330	86833	84788
venus	9	39816	39518	39039
venus	11	38243	38654	37882
venus	13	38191	39241	38006

Table I shows the number of incorrectly matched pixels by each method for 9 different image pairs with 3 different match window sizes. As can be seen from the table, we cannot claim that SSD is always better than SAD or vice versa. This means that one cannot use only one method all the time for best performance, which was argued before in this paper.

The number of incorrectly matched pixels are the lowest for most of the time with our new approach SSD+SAD, which shows its effectiveness. There are a few cases where our method does not produce the lowest number of incorrectly matched pixel. However, it never produces the highest

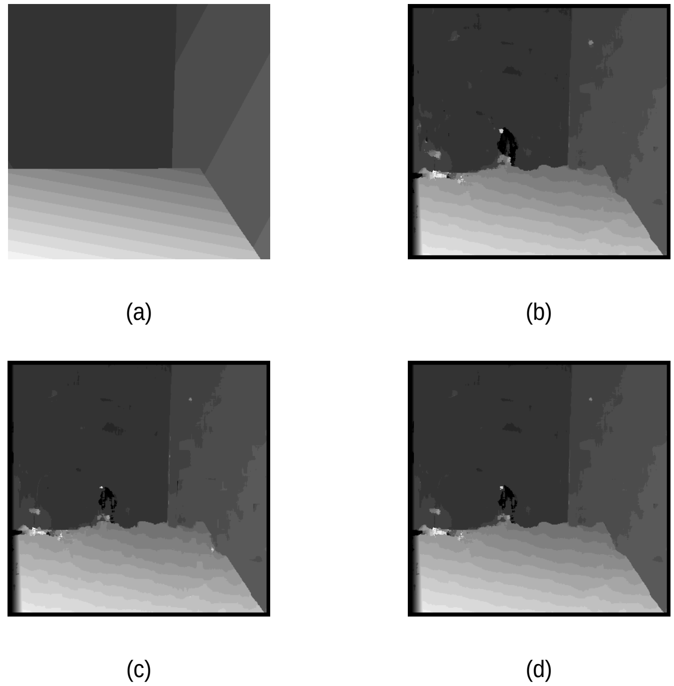

Fig. 4. (a) Ground truth disparity map for the bull image, (b) Depth map obtained by SAD function, number of error pixel is 41816, (c) Depth map obtained by SSD function, number of error pixel is 41199, (d)Depth map obtained by SAD+SSD function using our combination approach, number of error pixel is 40576. Percentage correction of SAD is 23% and percentage correction of SSD is 12%.

number of errors. In fact, if we average all the error values from all the experiments, we obtain the average error values shown in Table II. As it can be seen in Table II, our method (SSD+SAD) performs the best when we consider the average performance. This is due to the fact that our method always produces either the lowest number of incorrect matches or the second lowest, which cannot be said for either of SSD or SAD. We can conclude that it is always best to use our method for stereo correspondence because we know that SSD and SAD cannot perform at the same level if a number of images are considered. They can only perform better rarely for specific images due to random variability.

TABLE II

AVERAGE PIXEL MATCH ERRORS OF ALL EXPERIMENTS

Method	Average Errors
SSD	50370,52
SAD	50920,11
SSD+SAD	49656,89

We should also note that SSD performs second in the average error case, which might explain why it is more popular than SAD among the Computer Vision community.

It might seem that the differences between the performance of our method and the others are very small. Our method finds only about 700 correct pixels more on the average. However, it should be noted that our method is limited by the performance of SSD and SAD. We can find a correct pixel match only if one of the methods finds it correctly. For this reason, it is logical to calculate how many pixels are found correctly by SSD and incorrectly by SAD, and how many of them are corrected by our method SSD+SAD. The inverse process should also be calculated. Table III shows the results. The SSD+SAD method performance is bounded by the number of pixels that only one of the methods finds the correct results. The calculated values (Table IV) show that, our method can correct about 11% to 18% of the upperbound which is very encouraging. Considering that we do not use extra computational time in the correspondence, the obtained gain can be considered as valuable and effective.

V. CONCLUSIONS AND FUTURE WORK

Function optimization is a popular model feature enforcement technique in Computer Vision. Comparing and combining the results of different function optimizations are attractive and efficient ways of achieving more robust results. In this paper, we presented a new non-parametric statistical method that can assign a significance value to optimized function values. These values can be used to compare and combine different functions in a very efficient and effective way. The new method uses permutation tests without using lengthy computations to calculate the non-parametric distributions.

The ideas presented in this paper are applied in the stereo correspondence area to compare and combine the results obtained by SSD and SAD methods. The extensive

TABLE III

Image	Window Size	# of pixels correct by SSD incorrect by SAD	# of pixels correct by SAD incorrect by SSD	# of corrected in SSD	# of corrected in SAD
barn1	7	2952	2326	428(8%)	1054(20%)
barn1	9	2386	2540	912(19%)	758(15%)
barn1	11	2110	2700	1158 (24%)	568(12%)
barn2	7	4856	3048	-94 (-1%)	1714 (22%)
barn2	9	3863	2753	120 (2%)	1230 (19%)
barn2	11	3429	2527	73 (1%)	9754 (16%)
bull	7	5332	2684	-153(-2%)	2495(31%)
bull	9	4279	2465	283(4%)	2097(31%)
bull	11	3696	2443	470(8%)	1723(28%)
cones	9	5158	3235	1233(15%)	3156(38%)
cones	11	4486	3711	1568(19%)	2343(29%)
cones	13	3469	4470	1650(21%)	649(8%)
map	5	1734	1556	268(8%)	90(3%)
map	7	1314	1401	144(5%)	231(9%)
map	9	1003	1352	-4(0%)	345(15%)
poster	9	4233	2422	87(1%)	1898(29%)
poster	11	3912	2534	359(6%)	1757(27%)
poster	13	3544	2695	819(13%)	1668(27%)
sawtooth	7	2509	2820	692(13%)	381(7%)
sawtooth	9	1904	2914	1155(24%)	145(3%)
sawtooth	11	1771	3114	1346(28%)	3(0%)
teddy	7	5415	3713	753(8%)	2455(27%)
teddy	9	5269	4232	1470(15%)	2507(26%)
teddy	11	5095	4648	2045(21%)	2542(26%)
venus	9	3955	3657	479(6%)	777(10%)
venus	11	3099	3510	772(12%)	361(5%)
venus	13	2466	3516	1235(21%)	185(3%)

TABLE IV

AVERAGE CORRECTION PERCENTAGES FOR SSD AND SAD METHODS

Method	Average Correction
SSD	11%
SAD	18%

experiments showed that our combination method performs better than both methods on the average and it can correct the problems of these methods at a considerable level. We are very encouraged by these results and we plan to expand our method in many different directions including the possible solution to the classical Computer Vision regularization problem such as the weighting terms between the smoothness and the data energy in snakes or in graph cut algorithms.

We are also encouraged by the performance of the permutation test based vision systems and we plan to explore ways of using these techniques for the classification in the medical imaging and fault detection in industrial inspection.

REFERENCES

[1] S.Sarkar, K.L. Boyer, **Perceptual Organization in Computer Vision: A Review and a proposal for a Classificatory Structure**, IEEE Transactions on Systems, Man, and Cybernetics, vol.23, no.2, pp.382-399, Mar.1993.

[2] S. Sarkar, P. Soundararajan, **Supervised Learning of Large Perceptual Organization: Graph Spectral Partitioning and Learning Automata** , in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.5, pp.504-525, May 2000.

[3] A. Klaus, M.Sormann, K. Karner, **Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure**, The18th International Conference on Pattern Recognition (ICPR'06), IEEE, 2006.

[4] R.A. Fisher, **The Design of Experiment**, New York, 1935.

[5] Pitman, E. J. G.,**Significance tests which may be applied to samples from any population**, Royal Statistical Society Supplement, 1937; 4: 119-130 and 225-32 (parts I and II).

[6] E. Noreen, **Computer-intensive Methods for Testing Hypotheses**, Wiley, New York, 1989.

[7] M. Pagano, D. Trichler, **On obtaining permutation distribution in polynomial time, Journal of the American Statistical Association**, 1983.

[8] P.Golland, B. Fischl, **Permutation test for classification: Towards stastical significance in image based studies**, Information Processing in Medical Imaging, LNCS 2732, pp.330-341, Springer-Verlag Berlin Hidelberg 2003

[9] E.T. Bullmore, J. Suckling, S. Overmeyer, S. Hesketh, E. Toylor and M. Brammer **Global, Voxel, and Cluster Tests, by Theory and Permutation, for a Difference Between Two Groups of Structural MR Images of the Brain**, In IEEE Transactions on Medical Imaging, vol.18, no.1,pp. 32-42, 1999.

[10] T.E. Nichols and A.P. Holmes, **Nonparametric Permutation Test for Functional Neuroimaging: A Primer with Examples.**, Human Brain Mapping vol.15, pp.1-25, 2001.

[11] D. Scharstein and R. Szeliski. **High-accuracy stereo depth maps using structured light.** In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), volume 1, pages 195-202, Madison, WI, June 2003