J Math Imaging Vis (2013) 47:3–12
DOI 10.1007/s10851-012-0366-7

Tuning of Adaptive Weight Depth Map Generation Algorithms

Exploratory Data Analysis and Design of Computer Experiments (DOCE)

Diego Acosta · Iñigo Barandiaran · John Congote · Oscar Ruiz · Alejandro Hoyos · Manuel Graña

Published online: 6 July 2012

Abstract In depth map generation algorithms, parameters settings to yield an accurate disparity map estimation are usually chosen empirically or based on unplanned experiments. Algorithms' performance is measured based on the distance of the algorithm results vs. the Ground Truth by Middlebury's standards. This work shows a systematic statistical approach including exploratory data analyses on over 14000 images and designs of experiments using 31 depth maps to measure the relative influence of the parameters and to fine-tune them based on the number of bad pixels. The implemented methodology improves the performance of adaptive weight based dense depth map algorithms. As a result, the algorithm improves from 16.78 to 14.48 % bad pixels using a classical exploratory data analysis of over 14000 existing images, while using designs of computer experiments with 31 runs yielded an even better performance by lowering bad pixels from 16.78 to 13 %.

Keywords Stereo image processing · Parameter estimation · Depth map · Statistical design of computer experiments

D. Acosta
Grupo de Investigación DDP, Universidad EAFIT, Medellin, Colombia
e-mail: dacostam@eafit.edu.co

I. Barandiaran
Vicomtech Research Center, Donostia-San Sebastián, Spain
e-mail: ibarandiaran@vicomtech.org

J. Congote (✉) · O. Ruiz · A. Hoyos
Laboratorio CAD CAM CAE, Universidad EAFIT, Medellin, Colombia
e-mail: jcongote@eafit.edu.co

O. Ruiz
e-mail: oruiz@eafit.edu.co

A. Hoyos
e-mail: ahoyossi@eafit.edu.co

M. Graña
Dpto. CCIA, UPV-EHU, Donostia-San Sebastian, Spain
e-mail: ccpgrrom@gmail.com

1 Introduction

Depth map calculation deals with estimation of multiple object depths on a scene. It is useful for applications like vehicle navigation, automatic surveillance, aerial cartography, passive 3D scanning, industrial inspection, or 3D videoconferencing [1]. These maps are constructed by generating, at each pixel, an estimation of the distance from the camera to the object surface.

Disparity is commonly used to describe inverse depth in computer vision, and to measure the perceived spatial shift of a feature observed from close camera viewpoints. Stereo correspondence techniques often calculate a disparity function $d(x, y)$ relating target and reference images, so that the (x, y) coordinates of the disparity space match the pixel coordinates of the reference image. Stereo methods commonly use a pair of images taken with a known camera geometry to generate a dense disparity map with estimates at each pixel. This dense output is useful for applications requiring depth values even in difficult regions like occlusions and textureless areas. The ambiguity of matching pixels in these zones requires complex and expensive global image processing or statistical correlations using color and proximity measures in local support windows. The steps generally taken to compute the depth maps may include: (i) matching cost computation, (ii) cost or support aggregation, (iii) disparity computation or optimization, and (iv) disparity refinement.

In this article, Sect. 2 reviews the state-of-the-art. Section 3 describes our algorithm (filters, statistical analyses and experimental set-up). Section 4 discusses the results. Section 5 concludes the article.

2 Literature Review

Depth-map generation algorithms and filters use several user-specified parameters to generate a depth map from an image pair. The settings of these algorithms are heavily influenced by the evaluated data sets [3]. Published works usually report the settings used for their specific case studies without describing the procedure followed to fine-tune them [1, 4, 5], and some explicitly state the empirical nature of these values [6]. The variation of the output as a function of several settings on selected parameters is briefly discussed while not taking into account the effect of modifying them all simultaneously [3, 4, 7]. Reference [2] compares multiple stereo methods whose parameters are based on experiments. Only some parameters are tuned, without explaining the choices made. In the present article, we improve upon this work. In [8, 9], Depth Maps are generated from single images instead of image pairs.

2.1 Literature Review Conclusions

Used approaches in determining the settings of depth map algorithm parameters show all or some of the following shortcomings: (i) undocumented procedures for parameter setting, (ii) lack of planning when testing for the best settings, and (iii) failure to consider interactions of changing all parameters simultaneously.

As a response to these disadvantages, this article presents a methodology to fine-tune user-specified parameters on a depth map algorithm using a set of images from the adaptive weight implementation in [1]. Multiple settings are used and evaluated on all parameters to measure the contribution of each parameter to the output variance. A quantitative evaluation uses main effects plots and variance on multi-variate linear regression models to select the best combination of settings. Performance improves by setting new estimated values of user-specified parameters, allowing the algorithm to give much more accurate results on a rectified image pair.

Since it is not always feasible to have a large set of images available, a fractional factorial design of computer experiment (DOCE) with only eight runs is used to find out which parameters have a major influence on the images tested. To optimize the parameters and to have the lowest percentage of bad pixels a central composite DOCE with 23 runs is used with the most influential parameters found in the fractional factorial design. To the best of our knowledge the systematic and efficient application of DOCE in the field of depth maps generation has not been done yet.

3 Methodology

3.1 Image Processing

In adaptive weight algorithms [1, 4], a window is moved over each pixel on every image row, calculating a measurement based on the geometric proximity and color similarity of each pixel in the moving window to the pixel on its center. Pixels are matched on each row based on their support measurement with larger weights coming from similar pixel colors and closer pixels. The horizontal shift, or disparity, is recorded as the depth value, with higher values reflecting greater shifts and closer proximity to the camera.

The strength of grouping by color ($f_s(c_p, c_q)$) for pixels p and q is defined as the Euclidean distance between colors (Δc_{pq}) by Eq. (1). Similarly, grouping strength by distance ($f_p(g_p, g_q)$) is defined as the Euclidean distance between pixel image coordinates (Δg_{pq}) as per Eq. (2). γ_c and γ_p are adjustable settings used to scale the measured color delta, represented as *aw_col* in the study, and window size represented as *aw_win* respectively.

$$f_s(c_p, c_q) = \exp\left(-\frac{\Delta c_{pq}}{\gamma_c}\right) \tag{1}$$

$$f_p(g_p, g_q) = \exp\left(-\frac{\Delta g_{pq}}{\gamma_p}\right) \tag{2}$$

The matching cost between pixels shown in Eq. (3) is measured by aggregating raw matching costs, using the support weights defined by Eqs. (1) and (2), in support windows based on both the reference and target images.

$$E(p, \bar{p}_d) = \frac{\sum_{q \in N_p, \bar{q}_d \in N_{\bar{p}_d}} w(p,q) w(\bar{p}_d, \bar{q}_d) \sum_{c \in \{r,g,b\}} |I_c(q) - I_c(\bar{q}_d)|}{\sum_{q \in N_p, \bar{q}_d \in N_{\bar{p}_d}} w(p,q) w(\bar{p}_d, \bar{q}_d)} \tag{3}$$

where $w(p,q) = f_s(c_p, c_q) \cdot f_p(g_p, g_q)$, $\bar{p}_d$ and $\bar{q}_d$ are the target image pixels at disparity d corresponding to pixels p and q in the reference image, I_c is the intensity on channels red (r), green (g), and blue (b), and N_p is the window centered at p and containing all q pixels. The size of this movable window N is a derived parameter of (*aw_win*). Increasing the window size reduces the chance of bad matches at the expense of missing relevant scene features.

3.2 Post-Processing Filters

Algorithms based on correlations depend heavily on finding similar textures at corresponding points in both reference and target images. Bad matches happen more frequently in textureless regions, occluded zones, and areas with high

Table 1 Input and Output Variables of Depth Maps Generation Algorithms

Input Variables		
Parameter	*Description*	*Values*
Adaptive Weight [3]: Disparity estimation and pixel matching with γ_{aws}: similarity factor, and γ_{awg}: proximity factor related to the W_{AW} pixel size of the support window as user-adjustable parameters		
aw_win	Adaptive Weights Window Size	[1 3 5 7]
aw_col	Adaptive Weights Color Factor	[4 7 10 13 16 19]
Median: Smoothing and incorrect match removal with W_M: pixel size of the median window as user-adjustable parameter		
m_win	Median Window Size	[N/A 3 5]
Cross-check [8]: Validation of measurement per pixel with Δ_d: allowed disparity difference as adjustable parameter		
cc_disp	Cross-Check Disparity Delta	[N/A 0 1 2]
Bilateral [9]: Intensity and proximity weighted smoothing with edge preservation with γ_{bs}: similarity factor, and γ_{bg}: proximity factor related to the W_B pixel size of the bilateral window as user-adjustable parameters		
cb_win	Cross-Bilateral Window Size	[N/A 1 3 5 7]
cb_col	Cross-Bilateral Color Factor	[N/A 4 7 10 13 16 19]

Output Variables	
`rms_error_all`	Root Mean Square (RMS) disparity error (all pixels)
`rms_error_nonocc`	RMS disparity error (non-occluded pixels only)
`rms_error_occ`	RMS disparity error (occluded pixels only)
`rms_error_textured`	RMS disparity error (textured pixels only)
`rms_error_textureless`	RMS disparity error (textureless pixels only)
`rms_error_discont`	RMS disparity error (near depth discontinuities)
`bad_pixels_all`	Fraction of bad points (all pixels)
`bad_pixels_nonocc`	Fraction of bad points (non-occluded pixels only)
`bad_pixels_occ`	Fraction of bad points (occluded pixels only)
`bad_pixels_textured`	Fraction of bad points (textured pixels only)
`bad_pixels_textureless`	Fraction of bad points (textureless pixels only)
`bad_pixels_discont`	Fraction of bad points (near depth discontinuities)

variation in disparity, such as discontinuities. The winner-takes-it-all approach enforces uniqueness of matches only for the reference image so that points on the target image are matched more than once, creating the need to check the disparity estimates and to fill any gaps with information from neighboring pixels using post-processing filters like the ones discussed next (Table 1).

Median Filter (*m*) is widely used in digital image processing to smooth signals and to remove incorrect matches and holes by assigning neighboring disparities at the expense of edge preservation. The median filter provides a mechanism for reducing image noise, while preserving edges more effectively than a linear smoothing filter. It sorts the intensities of all q pixels on a window of size M and selects the median value as the new intensity of the p central pixel. The size M of the window is another of the user-specified parameters. *Cross-check Filter* (*cc*) performs twice the correlation by reversing the roles of the two images (reference and target) and considering valid only those matches having similar depth measures at corresponding points in both steps. The validity test is prone to fail in occluded areas where disparity estimates will be rejected. The allowed difference in disparities between reference and target images is one more adjustable parameter. *Bilateral Filter* (*cb*) is a non-iterative method of smoothing images while retaining edge detail. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels. The weighting for each pixel q is determined by the spatial distance from the center pixel p, as well as its relative difference in intensity, defined by Eq. (4).

$$O_p = \frac{\sum_{q \in W} f_s(q-p) g_i(I_q - I_p) I_q}{\sum_{q \in W} f_s(q-p) g_i(I_q - I_p)} \tag{4}$$

O_p is the output image, I the input image, W the weighting window, f_s the spatial weighting function, and g_i the intensity weighting function. The size of the window W is yet another parameter specified by the user.

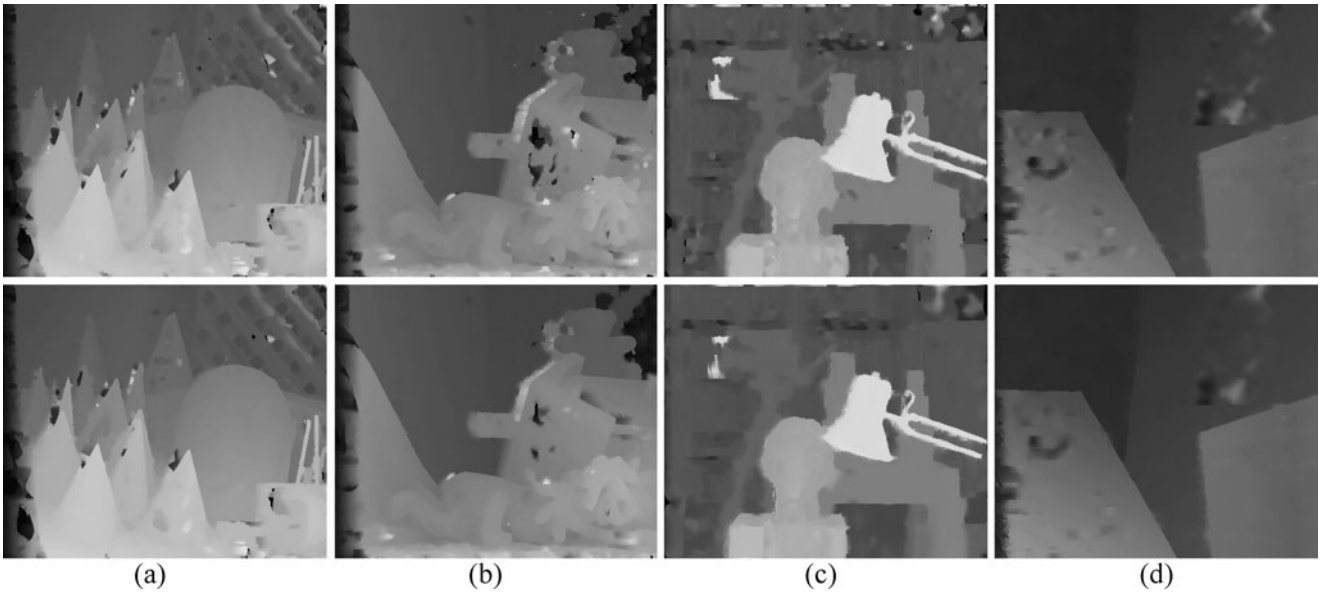

Fig. 1 Depth Map Comparison. Top: best initial, bottom: new settings. (**a**) Cones, (**b**) Teddy, (**c**) Tsukuba, and (**d**) Venus data set

3.3 Experimental Set-up

Our depth maps are calculated with an implementation developed for real time videoconferencing [1]. We use well-known rectified image sets: Cones from [2], Teddy and Venus from [10], and Tsukuba head and lamp from the University of Tsukuba. Our dataset consists of 14688 depth maps, 3672 for each data set, like the ones shown in Fig. 1.

Many recent stereo correspondence performance studies use the Middlebury Stereomatcher for their quantitative comparisons [3, 7, 11]. The evaluator code, sample scripts, and image data sets are available from the Middlebury stereo vision site, providing a flexible and standard platform for easy evaluation.

The online Middlebury Stereo Evaluation Table gives a visual indication of how well the methods perform with the proportion of bad pixels metric (`bad_pixels`) defined as the average of the proportion of bad pixels in the whole image (`bad_pixels_all`), the proportion of bad pixels in non-occluded regions (`bad_pixels_nonocc`), and the proportion of bad pixels in areas near depth discontinuities (`bad_pixels_discont`) in all data sets. A bad pixel represents a pixel where the estimated disparity is wrong with respect to a ground thruth disparity value.

3.4 Statistical Analyses

The user-specified input parameters and output accuracy data are statistically analyzed to correlate them (see Table 1). Box plots give insights on the influence of settings on a given response variable. Equation (5) relates $\hat{y}$ (predicted response) with x_i (input factors). β_0 and β_i are the coefficients fit by multi-variable linear regression. Constant Variance and Null Mean of Residuals help to validate the assumptions of the regression model. When those assumptions are not fulfilled, the model is modified [12]. The parameters are normalized to fit the range $(-1, 1)$ at their values shown in Table 1.

$$\hat{y} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon \tag{5}$$

Having a large data set (in this case 14688 images) to perform statistical analyses is not always feasible. DOCE is applied here to obtain an equivalently good model for the depth map, by having a much smaller number of runs. A 2^{6-3} fractional factorial DOCE with just eight runs allows to establish which ones of the parameters *aw_win*, *aw_colo*, *m_win*, *cc_disp*, *cb_win*, and *cb_col* are the most influential on the `bad_pixels` output by using a Daniel plot [13]. The parameters whose distribution cannot be considered as normal standard are statistically relevant in the fractional DOCE. Therefore, they are used to optimize the depth map generation algorithm.

A surface response central composite DOCE with 23 runs was performed afterward with *aw_win*, *aw_colo*, *m_win*, and *cb_win* as studied factors while keeping constant the remaining parameters (i.e., $cc\_disp = 2$ y $cb\_col = 13$) to yield a mathematical model of the form:

$$\hat{y} = \beta_0 + \sum_{i}^{k} \beta_i x_i + \sum_{ii}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j \tag{6}$$

where, as in Eq. (5), $\hat{y}$ is the predicted variable, x_i are the parameters, and β_0, β_i, β_{ii} and β_{ij} are constants adjusted by

minimum least squares regression. Data from DOCE was analysed with the software for statistical computing R with Bayes Screening and Model Discrimination -BsMD- and Response Surface Method -rsm- add-on packages [14].

4 Results and Discussion

4.1 Selection of Input Variables for Mathematical Model

Response variables for depth map generation algorithms are shown with their meaning in Table 1. Pearson multiple correlation coefficients for the response variables shown in Table 2 evidences that `bad_pixels_all` is strongly correlated to the remaining response variables. This means that all response variables follow a similar trend as `bad_pixels_all` and that modeling `bad_pixels_all` is sufficient to reach statistically sound results for depth map generation algorithms optimization.

On the other hand, low Pearson coefficients for the input variables indicate that those variables are independent, that there is no co-linearity among them and that each independent variable must be included in the exploratory analysis.

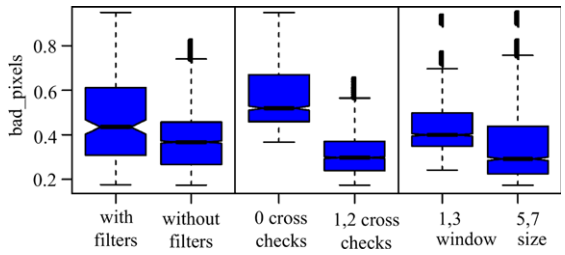

Fig. 2 Box Plots for Input Variable Analysis

4.2 Exploratory Data Analysis

Box plots analyses of `bad_pixels` presented in Fig. 2 shows lower output values from using filters, relaxed cross-check disparity delta values, large adaptive weight window sizes, and large adaptive weight color factor values. The median window size, bilateral window size, and bilateral window color values do not show a significant influence on the output at the studied levels.

The influence of the parameters is also shown by the value of the slopes of the main effects plots in Fig. 3 and confirms the behavior found with the analysis of variance (ANOVA) of the multi-variate linear regression model. The optimal settings from this analysis (i.e., $aw\_win = 9$, $aw\_col = 22$, $m\_win = 5$, $cc\_disp = 1$, $cb\_win = 3$ and $cb\_col = 4$) to minimize `bad_pixels` yields a result of 14.48 %.

4.3 Multi-variate Linear Regression Model

The analysis of variance on a multi-variate linear regression (MVLR) over all data sets using the most parsimonious model quantifies the parameters with the most influence as shown in Table 3. The most significant input variable is *cc_disp*, since it accounts for a [33–50 %] of the variance in every case.

Interactions and higher order terms are included on the multi-variate linear regression models to improve the goodness of fit. Reducing the number of input images per dataset from 3456 to 1526 by excluding the worst performing cases ($cc\_disp = 0$, $aw\_col = 4$ and $aw\_col = 7$), using a cubic model with interactions yields a very good multiple correlation coefficient of $R^2 = 99.05$ %. However, for the model

Table 2 Pearson correlation coefficient for the evaluator outputs over all data sets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) `bad_pixels`	1.00	0.81	0.82	0.59	0.83	0.77	0.84	1.00	1.00	0.86	1.00	0.95	0.99
(2) `rms_error_all`	0.81	1.00	1.00	0.69	1.00	0.98	0.99	0.82	0.82	0.64	0.85	0.70	0.79
(3) `rms_error_nonocc`	0.82	1.00	1.00	0.71	1.00	0.98	0.99	0.83	0.82	0.67	0.85	0.71	0.80
(4) `rms_error_occ`	0.59	0.69	0.71	1.00	0.70	0.77	0.74	0.62	0.61	0.68	0.61	0.63	0.53
(5) `rms_error_textured`	0.83	1.00	1.00	0.70	1.00	0.98	0.99	0.83	0.83	0.67	0.86	0.72	0.81
(6) `rms_error_textureless`	0.77	0.98	0.98	0.77	0.98	1.00	0.98	0.78	0.78	0.64	0.80	0.68	0.73
(7) `rms_error_discont`	0.84	0.99	0.99	0.74	0.99	0.98	1.00	0.85	0.84	0.67	0.87	0.73	0.82
(8) `bad_pixels_all`	1.00	0.82	0.83	0.62	0.83	0.78	0.85	1.00	1.00	0.85	1.00	0.96	0.98
(9) `bad_pixels_nonocc`	1.00	0.82	0.82	0.61	0.83	0.78	0.84	1.00	1.00	0.85	1.00	0.96	0.98
(10) `bad_pixels_occ`	0.86	0.64	0.67	0.68	0.67	0.64	0.67	0.85	0.85	1.00	0.83	0.87	0.86
(11) `bad_pixels_textured`	1.00	0.85	0.85	0.61	0.86	0.80	0.87	1.00	1.00	0.83	1.00	0.93	0.99
(12) `bad_pixels_textureless`	0.95	0.70	0.71	0.63	0.72	0.68	0.73	0.96	0.96	0.87	0.93	1.00	0.93
(13) `bad_pixels_discont`	0.99	0.79	0.80	0.53	0.81	0.73	0.82	0.98	0.98	0.86	0.99	0.93	1.00

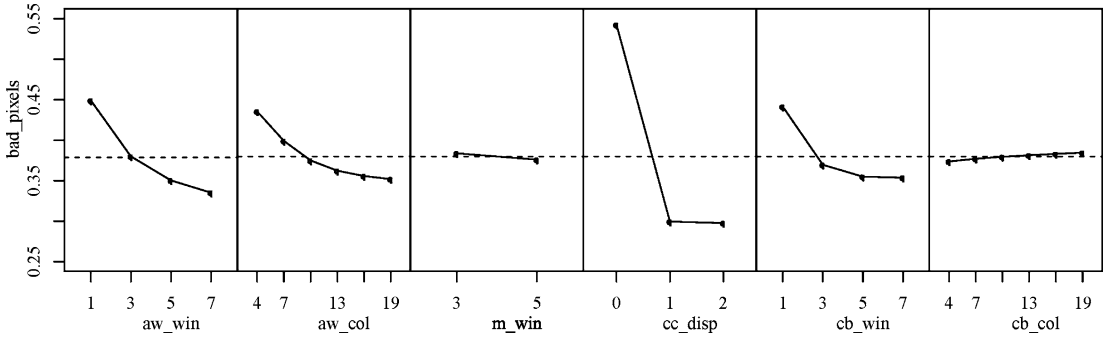

Fig. 3 Main Effects Plots of each factor level for all data sets. Steeper slopes relate to bigger influence on the variance of the `bad_pixels` output measurement

selected the residuals distribution is not normal even after transforming the response variable and removing large residuals values. Another constraint for the statistical analyses is that any outliers from the data set can not be excluded. Nonetheless, improved algorithm performance settings are found using the model to obtain lower `bad_pixels` values comparable to the ones obtained through the exploratory data analysis (14.66 % vs. 14.48 %).

In summary, the most noticeable influence on the output variable comes from having a relaxed cross-check filter, accounting for nearly half the response variance in all the study data sets. Window size is the next most influential factor, followed by color factor, and finally window size on the bilateral filter. Increasing the window size on the main algorithm yields better overall results at the expense of longer running times and some foreground loss of sharpness, while the support weights on each pixel have the chance of becoming more distinct and potentially reduce disparity mismatches. Increasing the color factor on the main algorithm allows better results by reducing the color differences, and slightly compensating minor variations in intensity from different viewpoints.

A small median smoothing filter window size is faster than a larger one, while still having a similar accuracy. Low settings on both the window size and the color factor on the bilateral filter seem to work best for a good trade-off between performance and accuracy.

The optimal settings in the original data set are presented in Table 4 along with the proposed settings. *Low settings* comprise the depth maps with all their parameter settings at each of their minimum tested values yielding 67.62 % `bad_pixels`. *High settings* relates to depth maps with all their parameter settings at each of their maximum tested values yielding 19.84 % `bad_pixels`. *Best initial* are the most accurate depth maps from the study data set yielding 16.78 % `bad_pixels`. *Exploratory analysis* corresponds to the settings determined using the exploratory data analysis based on box plots and main effects plots yielding 14.48 % `bad_pixels`. *MVLR optimization* is the optimization of the classical data analysis based on multi-variate

Table 3 Linear model ANOVA with the contribution to the sum of squared errors (SSE) of `bad_pixels`

Data set	*cc_disp*	*aw_win*	*aw_col*	*cb_win*
Cones	34.35 %	14.46 %	17.47 %	–
Teddy	41.25 %	13.75 %	8.10 %	–
Tsukuba	50.25 %	–	–	7.16 %
Venus	47.35 %	9.42 %	–	5.62 %
All	47.01 %	8.11 %	–	–

Table 4 Model comparison. Average `bad_pixels` values over all data sets and their parameter settings

Run Type	`bad_pixels`	*aw_win*	*aw_col*	*m_win*	*cc_disp*	*cb_win*	*cb_col*
Low Settings	67.62 %	1	4	3	0	1	4
High Settings	19.84 %	7	19	5	2	7	19
Best Initial	16.78 %	7	19	5	1	3	4
Exploratory analysis	14.48 %	9	22	5	1	3	4
MVLR optimization	14.66 %	11	22	5	3	3	18
Best Treatment for Fractional Factorial DOCE	14.72 %	10	25	3	3	1	3
Best Treatment for CCD DOCE	13.05 %	7	14	3	4	1	13

linear regression model, nested models, and ANOVA yielding 14.66 % `bad_pixels`.

The exploratory analysis estimation and the MVLR optimization tend to converge at similar lower `bad_pixels` values using the same image data set. The best initial and improved depth map outputs are shown in Fig. 1. The best runs for fractional factorial and central composite DOCEs lower the value of the `bad_pixels` variable to 14.72 % and 13.05 %, respectively. Notice that to achieve these results only 31 depth maps are needed (DOCE) as opposed to analyzing over 14000 depth maps (Exploratory Analysis).

4.4 Depth-Map Optimization by Design of Computer Experiments (DOCE)

2^{6-3} *Fractional Factorial Design of Experiment* The goal of this type of design of experiment is to screen the statistically most significant parameters. Details on how to set up the runs are discussed in [12]. The design matrix describing all experimental runs can be set so that the high and low levels for each parameter are chosen by assigning them the maximum and minimum values allowed by the algorithm respectively. This was done for all of the parameters but for *m_win* (i.e., it was set at the levels 3 and 5), to avoid bias from the results and conclusions obtained from the exploratory and multivariate regression analysis. The results for this DOCE range from 14.72 and 72.17 % bad pixels for all images which is quite promising because already with only eight runs a set of parameters values that is very close to the optimum obtained by exploratory analysis of 14.48 % bad pixels and the multivariate linear regression analysis of 14.66 % on the 14688 data points is delivered. The alias for the parameters and Daniel plot showing the most relevant ones are shown in Fig. 4.

Daniel's plot indicate that the most influential parameters are *cc_disp*, *aw_win* and *cb_win* which deviate the most from the normal distribution curve. These parameters and *m_win* at levels 0, 3 and 5 are used for the surface response methodology central composite design of experiment that follows.

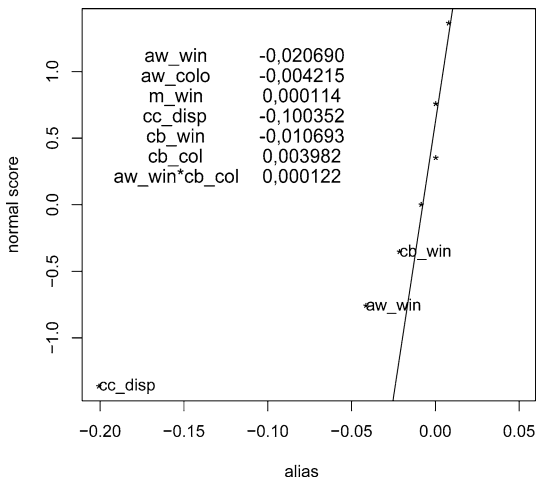

Fig. 4 Daniel Plot for determining the significance of input variables

Central Composite Design of Experiment To further optimize the depth maps generation algorithm a central composite design of experiment is used. As with the fractional factorial design of experiment, the best run with 13.05 % bad pixels is obtained amongst the 23 treatments which surpasses the results obtained thus far. The outputs from R using the rsm package at the levels tested for each parameter are shown in Table 5.

As it can be seen the second order model depicted before in Eq. (6) fits very well the data as indicated by the multiple correlation coefficient 0.9695. The most significant variables include *aw_win*, *aw_col*, *m_win*, *cb_win*, aw_win^2, aw_col^2, and m_win^2. Nonetheless, the complete model with all coefficients is used to draw the contour plots shown later. The rsm package also allows to detect stationary points. In this case the stationary point detected is a saddle point because one of the eigen-values is negative while the remaining ones are positive

Graphically the iso-lines for `bad_pixels_all` are seen in slices by looking at two parameters simultaneously for the analysis while keeping the remaining ones constant as shown in Fig. 5. The graphs allow to see that the stationary point does indicate a local minimum when analyzing for *aw_win* and *aw_col*. With *m_win* though, the graph indicates that a saddle is detected and that it is better to use values not in the $1.5 < m\_win < 3.5$ interval (which is physically imposible). For *cb_win* the stationary point apparently corresponds to a minimum. The settings for the stationary point closer to what rsm's package detects are $aw\_size = 7$, $aw\_col = 14$, $m\_size = 3$, $cc\_disp = 2$, $cb\_size = 21$ and $cb\_col = 13$ and this yields 26 % `bad_pixels_all` leading to conclude that the best treatment for the rsm yielding 13.05 % of `bad_pixels_all` is the local minimum optimum at the settings shown on Table 4.

5 Conclusions and Future Work

Previously published material in [15] showed how Exploratory Analysis, applied on over 14000 images, allowed the sub-optimal tuning of the parameters for Disparity Estimation algorithms, lowering the percentage of bad pixels from 16.78 % (manual tuning) to 14.48 %. The present work shows how to use DOCE to optimize the tuning, by running a dramatically smaller sample (31 experiments). The result of applying DOCE allowed to reach 13.05 % of bad pixels, without the need of Exploratory Analysis. Using DOCE reduces the number of depth maps needed to carry out the

Table 5 Summary of RSM Central Composite DOCE

Parameters	Levels				
aw_win	1	4	7	10	13
aw_colo	3	8.5	14	19.5	25
m_win	0	3	5		
cb_win	1	4	7	10	13
cb_colo	13				
cc_disp	2				

Call: rsm(formula = `bad_pixels_all` ~ SO(*aw_win*, *aw_colo*, *m_win*, *cb_win*))

Coefficients	Estimate	Std. Error	t-value	$p > \|t\|$	Signif.
(Intercept)	1.634	2.49×10^{-1}	6.561	0.00018	***
aw_win	-5.25×10^{-1}	8.03×10^{-2}	−6.538	0.00018	***
aw_colo	-1.9×10^{-1}	4.78×10^{-2}	−3.971	0.00411	**
m_win	1.606	4.22×10^{-1}	3.802	0.00522	**
cb_win	-3.96×10^{-2}	8.03×10^{-2}	−0.493	0.63495	
aw_win: *aw_colo*	4.15×10^{-5}	1.59×10^{-4}	0.26	0.80128	
aw_win: *m_win*	-6.13×10^{-5}	7.01×10^{-4}	−0.087	0.93243	
aw_win: *cb_win*	1.73×10^{-4}	2.92×10^{-4}	0.592	0.56990	
aw_colo: *m_win*	3.01×10^{-4}	3.82×10^{-4}	0.788	0.45339	
aw_colo: *cb_win*	5.33×10^{-4}	1.59×10^{-4}	3.347	0.01013	*
m_win: *cb_win*	5.56×10^{-4}	7.01×10^{-4}	0.793	0.45083	
*aw_win*ˆ2	3.73×10^{-2}	5.73×10^{-3}	6.508	0.00019	***
*aw_colo*ˆ2	6.44×10^{-3}	1.70×10^{-3}	3.78	0.00539	**
*m_win*ˆ2	-3.25×10^{-1}	8.45×10^{-2}	−3.846	0.00490	**
*cb_win*ˆ2	7.19×10^{-4}	5.73×10^{-3}	0.126	0.90314	

Signifficance codes: 0^{***}, 0.001^{**}, 0.01^{*}, 0.05, 0.1, 1

Residual standard error: 0.04208 on 8 degrees of freedom

Multiple R-squared: 0.9695, Adjusted R-squared: 0.916

F-statistic: 18.13 on 14 and 8 DF, p-value: 0.0001577

Stationary point at response surface		Eigen-values	
aw_win	6.987	λ_1	0.0373
aw_colo	13.788	λ_2	0.0064
m_win	2.495	λ_3	0.0007
cb_win	20.615	λ_4	−0.3249

study when a large image database is not available. The DOCE methodology itself is independent of the particular algorithms used to generate the disparity maps and it can be used whenever a systematic tunning of process parameters is required.

An improvement from 16.78 % (manual tuning) to 13.05 % in the `bad_pixels_all` variable might seem negligible at first glance. However, such figures imply a jump of the optimized algorithm of almost 10 positions in the Middlebury Stereo Evaluation ranking. It must be noticed that many algorithms competing in such a rank could benefit from the systematic tunning presented here.

A Surface Reconstruction application with DOCE uses the optimal tuning of disparity maps between two stereoscopic images scanning a scene. The disparity between the images, in turn, allows the triangulation of the 3D points on the surface of objects in the scene. This point cloud is an input to surface reconstruction algorithms. This process is

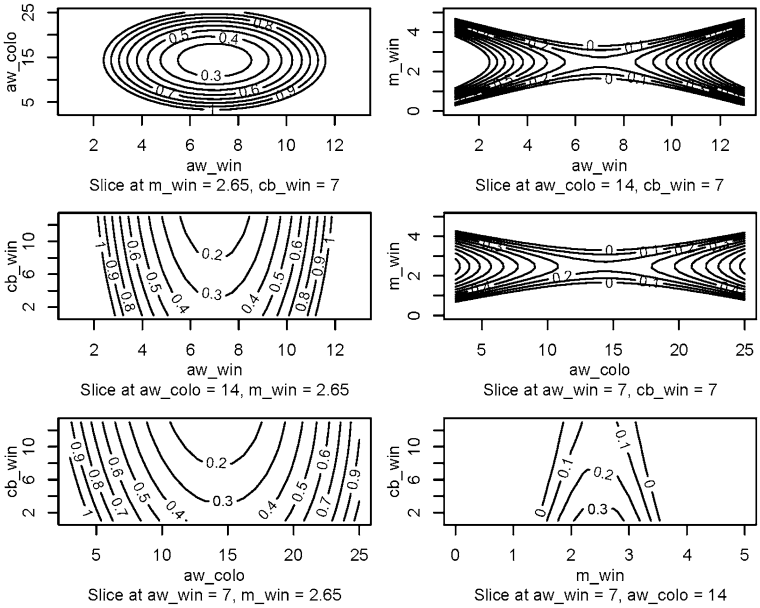

Fig. 5 Contour Plots for Central Composite DOCE

discussed in detail in [1]. DOCE applications in other domains are indeed possible.

Acknowledgements This work has been partially supported by the Spanish Administration Agency CDTI under project CENIT-VISION 2007-1007, the Colombian Administrative Department of Science, Technology, and Innovation; and the Colombian National Learning Service (COLCIENCIAS-SENA) grant No. 1216-479-22001.

References

1. Congote, J., Barandiaran, I., Barandiaran, J., Montserrat, T., Quelen, J., Ferrán, C., Mindan, P., Mur, O., Tarrés, F., Ruiz, O.: Real-time depth map generation architecture for 3D videoconferencing. In: 3DTV-CON, 2010, pp. 1–4 (2010)
2. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**(1–3), 7–42 (2002)
3. Gong, M., Yang, R., Wang, L., Gong, M.: A performance study on different cost aggregation approaches used in real-time stereo matching. Int. J. Comput. Vis. **75**, 283–296 (2007)
4. Yoon, K., Kweon, I.: Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 650 (2006)
5. Gu, Z., Su, X., Liu, Y., Zhang, Q.: Local stereo matching with adaptive support-weight, rank transform and disparity calibration. Pattern Recognit. Lett. **29**, 1230–1235 (2008)
6. Hosni, A., Bleyer, M., Gelautz, M., Rhemann, C.: Local stereo matching using geodesic support weights. In: 16th IEEE International Conference on Image Processing (ICIP), pp. 2093–2096 (2009)
7. Wang, L., Gong, M., Gong, M., Yang, R.: How far can we go with local optimization in real-time stereo matching. In: Proceeding 3DPVT'06 Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), pp. 129–136 (2006)
8. Battiato, S., Curti, S., La Cascia, M., Scordato, E., Tortora, M.: Depth map generation by image classification. In: Proceeding of SPIE Electronic Imaging 2004. Three-Dimensional Image Capture and Applications VI, vol. 5302-13 (2004)
9. Battiato, S., Capra, A., Curti, S., La Cascia, M.: 3D stereoscopic pairs by depth-map image generation. In: IEEE 3DPVT'04, 2nd Int. Symp. on 3D Data Processing Visualization & Transmission, pp. 124–131 (2004)
10. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: CVPR IEEE, vol. 1, pp. 195–202 (2003)
11. Tombari, F., Mattoccia, S., Di Stefano, L., Addimanda, E.: Classification and evaluation of cost aggregation methods for stereo correspondence. In: CVPR IEEE, pp. 1–8 (2008)
12. Montgomery, D.C.: Design and Analysis of Experiments (2010). ISBN:978-0-470-88606-9
13. Daniel, C.: Use of half-normal plots in interpreting factorial two-level experiments. Technometrics **1**(4), 311–341 (1959)
14. Lenth, R.V.: Response-Surface methods in R, using rsm. J. Stat. Softw. **32**(7), 1–17 (2009)
15. Hoyos, A., Congote, J., Barandiaran, I., Acosta, D., Ruiz, O.: Statistical tuning of adaptive-weight depth map algorithm. In: CAIP, pp. 563–572 (2011)

Diego Acosta obtained a degree of Chemical Engineer from the Universidad Pontificia Bolivariana (UPB) Medellín, Colombia and a M.Sc. and Ph.D. from University of Oklahoma, Norman, USA. Dr. Eng. Acosta worked in "Ashland Chemical Company" (2001–2002) and in "Xerox, Oklahoma City" (2000–2001) as process engineer. He was Thesis Coordinator of the Chemical Engineering Department at the UPB (1998–1999) and superintendence assistant of the recovery and power plant of Smurfit Cartón at Cali, Colombia (1993–1995). Dr. Eng. Acosta is currently Associate Professor at EAFIT University, Medellín, Colombia since 2007. His research interests are Statistics and Design of Experiments applied to the Process Engineering. Prof. Acosta supervises the courses Statistics, Design of Experiments, Desing in Process Engineering, Mass Transfer Laboratory, and Process Optimization.

Iñigo Barandiaran studied Computer Engineering in the Basque Country University (http://www.ehu.es) between 1995 and 2001. Between 2001 and 2003, he was a scholarship holder in the department of Computer Science and Artificial Intelligence of the Universidad del Pais Vasco and attended a Pre-Doc program at that University. He has been a researcher in the VICOMTech-IK4 fundation (http://www.vicomtech.org), in the Bio-Medical Applications area since 2003. His main research topic is focused on the development of feature extraction and matching techniques applied in optical tracking.

John Congote (Medellín, Colombia) is a Computer Science student at EAFIT University, and research assistant in the CAD/CAM/CAE Laboratory EAFIT since 2007. John has successfully participated in several programming competitions since 1996. John Congote obtained the M.Sc. in Informatics with Honors in June 2009 with his work in the Institute for Visual Communication Technologies "Vicomtech" in San Sebastian, Spain (2008–2009). His M.Sc. Thesis researched in Computational Geometry applied to CAD and Computer Vision. John Congote is currently Doctoral Student of the CAD CAM CAE Laboratory, based in Vicomtech for the duration of his Ph.D. John's main interest areas are algorithms, computational geometry, computer graphics and programming languages.

Oscar Ruiz was born in 1961 in Tunja, Colombia. He obtained B.Sc. degrees in Mechanical Eng. (1983) and Computer Science (1987) at Los Andes University, Bogota, Colombia, a M.Sc. degree with emphasis in CAM (1991) and a Ph.D. with emphasis in CAD (1995) from the Mechanical & Industrial Eng. Dept. of University of Illinois at Urbana, Champaign, USA. Dr. Ruiz has held Visiting Researcher positions at Ford Motor Co. (Dearborn, USA. 1993 and 1995), Fraunhofer Inst. Graphische Datenverarbeitung (Darsmstad, Germany 1999 and 2001), University of Vigo (1999 and 2002), Max Planck Institute for Informatik (2004) and Purdue University (2009). In 1996 Dr. Ruiz was appointed as Faculty of the Mechanical Eng. and Computer Science Depts. at EAFIT University, Medellín, Colombia, and has ever since the coordinator of the Laboratory for Interdisciplinary Research on CAD/CAM/CAE. Dr. Ruiz' interests are Computer Aided Geometric Design, Geometric Reasoning and Applied Computational Geometry.

Alejandro Hoyos was born in 1978 in Medellín, Colombia. He is an undergraduate student of Mechanical Engineering at EAFIT University. He was a research and teaching assistant in the CAD CAM CAE Laboratory-EAFIT since 2009 under supervision of Prof. Dr. Eng. Oscar Ruiz, and took Mechanical Engineering courses at Concordia University in Montreal, Canada during two terms after being selected in the International Student Exchange Program at EAFIT. He was a Six Sigma Black Belt while working at Andercol in Medellín, Colombia, and is scheduled to obtain his Diploma in Mechanical Engineering from EAFIT University with his Graduation Project under the advisory of Professors Oscar Ruiz and Diego Acosta on statistical analysis of input factors on an adaptive weight depth map algorithm.

Manuel Graña is a full professor of Department of Computer Science and Artificial Intelligence of the Universidad del Pais Vasco Research interests: control, distributed and embedded systems, image processing, machine learning, bio-inspired computing, robotics, computer systems performance modeling, social network modeling Achievements: coauthored over 80 papers in international journals, over 150 papers in conferences, over 10 books edited. Directed over 10 research projects funded by the Spanish government, and has been advisor of some 20 Ph.D. Thesis students.