Similarity measures for image matching despite occlusions in stereo vision

Sylvie Chambon, Alain Crouzil

Similarity measures for image matching despite occlusions in stereo vision

Sylvie Chambon[1,2] (corresponding author) and Alain Crouzil[2]

[1]*LCPC, MACS, AI, route de Pornic, 44341 Bouguenais Cedex, France*
Tel: +33 2 40 84 56 13
[2]*UPS, IRIT, TCI, 118 route de Narbonne, 31062 Toulouse Cedex 9, France*

Abstract

In the context of computer vision, matching can be done with similarity measures. This paper presents the classification of these measures into five families. In addition, eighteen measures based on robust statistics, previously proposed [1] in order to deal with the problem of occlusions, are studied and compared to the state of the art. A new evaluation protocol and new analyses are proposed and the results highlight the most efficient measures, first, near occlusions, the smooth median powered deviation, and second, near discontinuities, a non-parametric transform-based measure, CENSUS.

Key words: stereo vision, image matching, similarity, correlation, occlusions.

1 Introduction

In computer vision, similarity metrics are widely employed: in image registration [2], pattern recognition [3], movement analysis [4], object tracking [5], video analysis [6] and stereo matching [7]. Consequently, many publications introduce new similarity measures and some papers give a review of these measures. The most popular is the taxonomy of Aschwanden and Guggenbül [7] but we can also mention taxonomies of ordinal measures [8], robust measures for matching [9], for

Email address: chambon@lcpc (Sylvie Chambon[1,2] (corresponding author) and Alain Crouzil[2]).

registration [2] and for classification [10]. For stereo matching, correlation-based methods are popular, because the implementation is simple, the execution time is low and their efficiency has been demonstrated [11]. Matching elements can be pixels or more complex features [12] such as edges or corners. In this paper, we deal with dense pixel matching. We consider that a correlation measure evaluates the similarity between two data sets: the grey levels of two pixels and their neighbourhoods. Even if some of the correlation measures (classical, derivative-based and non-parametric transform-based) have been studied and compared [13], the choice of one measure is difficult. So, it seems important to give a new analysis of the existing correlation measures.

Stereo matching is difficult because of: intensity distortions, noise, untextured areas, foreshortening, perspective effects and occlusions. Intensity distortions and noise have been investigated [7], whereas untextured areas and foreshortening cannot be overcome with correlation. With perspective effects, fine correlation [5] can be used. While solving the occlusion problem, adaptive windows [14], multiple-window methods [15], support-weight approaches [11] or robust measures [9] can be used. Here, we are particularly concerned with occlusions. In a scene, depth discontinuities induce the occlusion problem because it is difficult to match a pixel whose neighbours have a different depth. One solution is to consider as outliers the pixels having a different depth from the pixel being studied. The tools of robust statistics are insensitive to outliers and, consequently, we propose to introduce robust statistics-based measures. The aim of this work is to evaluate the correlation measure and not to evaluate the window strategy for the estimation of the similarity and we do not consider this kind of approach [11,14].

After giving some notations, the commonly used correlation measures are presented. Then, eighteen robust correlation measures are proposed. Finally, an evalu-

ation protocol is presented and the results are discussed.

2 Notations

We propose to classify the measures into five families, with measures based on: cross correlation, classical statistics, derivative images, non-parametric statistics and robust statistics, see notations in Table 1.

I_w	The images with $w \in \{l, r\}$ (left and right).
$I_{\max}$	The maximal grey level in the image I [a].
$I_w^{i,j}$, $\mathbf{p}_w^{i,j}$	The grey level of the pixel $\mathbf{p}_w^{i,j}$ of coordinates (i, j) in image I_w is $I_w^{i,j}$. Moreover, $\mathbf{p}_r^{v,w}$ is the correspondant pixel of $\mathbf{p}_l^{i,j}$.
N_f, N_v, N_h	The number of pixels in the correlation window is denoted by: $N_f = (2N_v + 1) \times (2N_h + 1)$, N_v, $N_h \in \mathbb{N}^*$.
$\mathbf{f}_w$	This vector contains grey levels of pixels in the correlation window (in image I_w): $\mathbf{f}_w = (\cdots I_w^{i+p,j+q} \cdots)^T = (\cdots f_w^k \cdots)^T$ where ^{T} is the matrix transposition operator, $p \in [-N_v; N_v]$, $q \in [-N_h; N_h]$.
f_w^k	The element k of vector $\mathbf{f}_w$.
$\overline{\mathbf{f}_w}$, $\mathrm{m}(\mathbf{f}_w)$	The vector of means $\overline{\mathbf{f}_w}$ contains N_f columns and is defined by: $\overline{\mathbf{f}_w} = (\mathrm{m}(\mathbf{f}_w) \ \cdots \ \mathrm{m}(\mathbf{f}_w))^T$ with $\mathrm{m}(\mathbf{f}_w) = \frac{1}{N_f} \sum_{k=0}^{N_f-1} f_w^k$.
L_P	The L_P norms are: $\|\mathbf{f}_w\|_P = \left(\sum_{k=0}^{N_f-1} \lvert f_w^k \rvert^P \right)^{1/P}$ with $P \in \mathbb{N}^*$. The Euclidean norm is given by: $\|\mathbf{f}_w\| = \|\mathbf{f}_w\|_2$.

[a] In this paper $I_{\max} \in [0; 255]$.

Table 1
Notations used for the description of the measures.

Tables 2 to 9 present for each measure, the following details: the name of the measure, the abbreviation of the measure, the formulae, a lower bound and an upper bound of the interval of variation (VARIATION). For each measure, with $a,\ b \in \mathbb{R}^*$ [1], and, $c,\ d \in \mathbb{R}$, we define the invariance property:

[1] Like $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$

Gain	:	$\mathrm{M}(a\mathbf{f}_l, b\mathbf{f}_r) = \mathrm{M}(\mathbf{f}_l, \mathbf{f}_r)$;
Bias	:	$\mathrm{M}(\mathbf{f}_l + c, \mathbf{f}_r + d) = \mathrm{M}(\mathbf{f}_l, \mathbf{f}_r)$;
Gain and bias	:	$\mathrm{M}(a\mathbf{f}_l + c, b\mathbf{f}_r + d) = \mathrm{M}(\mathbf{f}_l, \mathbf{f}_r)$.

In the following description, when no explicit reference is given, the reader should consult Aschwanden and Guggenbül [7].

3 Cross correlation-based measures

The CROSS family (Table 2) is based on the scalar product. The cross correlation:

$$\mathrm{CC}(\mathbf{f}_l, \mathbf{f}_r) = \mathbf{f}_l \cdot \mathbf{f}_r \tag{1}$$

can be used only if the vectors $\mathbf{f}_w$ are normalised. This normalisation brings gain invariance [16] and leads to the Normalised Cross Correlation, NCC (similarity measure). The centred version, called the Zero mean Normalised Cross Correlation, ZNCC, gain and bias invariant, is also known as an estimation of the Pearson product-moment correlation coefficient. It is more efficient than NCC when there is a linear relationship between the two sets of grey levels to be compared. The Moravec [17] similarity measure, MOR, proposed for binary images, uses a different normalisation which is faster to compute than the normalisation of ZNCC. It has been proposed to solve the problems of ZNCC, when the denominator is equal to zero, but, consequently, MOR is sensitive to contrast changes and only bias invariant.

NAME	ABBREVIATION	DEFINITION	VARIATION
Normalised Cross Correlation	NCC	$\mathbf{f}_l \cdot \mathbf{f}_r / \|\mathbf{f}_l\| \|\mathbf{f}_r\|$	$[0;1]$
Zero mean Normalised Cross Correlation	ZNCC	$\mathrm{NCC}(\mathbf{f}_l - \overline{\mathbf{f}_l}, \mathbf{f}_r - \overline{\mathbf{f}_r})$	$[-1;1]$
Moravec [17]	MOR	$\frac{2(\mathbf{f}_l - \overline{\mathbf{f}_l}) \cdot (\mathbf{f}_r - \overline{\mathbf{f}_r})}{\|\mathbf{f}_l - \overline{\mathbf{f}_l}\|^2 + \|\mathbf{f}_r - \overline{\mathbf{f}_r}\|^2}$	$[-1;1]$

Table 2
CROSS family.

4 Classical statistics-based measures

The CLASSICAL family (Table 3) contains classical statistics-based dissimilarities: the distances, the locally scaled distances [7], the variances [18] and a fourth-order statistics-based measure [19].

Distances – The principle behind the use of a distance in order to quantify the similarity between two sets of grey levels is to consider them as two points in $\mathbb{R}^{N_f}$ and to estimate how distant they are. In other words, it consists in calculating the L_P norms of the vector of the grey level differences [20,21]. The main measures are the Sum of Absolute Differences, SAD (L_1 norm), the Sum of Squared Differences, SSD (L_2 norm), and the Kolmogorov-Smirnov distance, D_∞ (L_∞ norm). These measures can be centred to be invariant to bias, leading to Zero mean Distances, ZD_P. The well-known centred measures are the Zero mean Sum of Absolute Differences, ZSAD (ZD_1), and the Zero mean Sum of Squared Differences, ZSSD (ZD_2). These measures can also be normalised, leading to Normalised Distances, ND_P, including the Normalised Sum of Squared Differences, NSSD (ND_2), and, centred and normalised, giving the Zero mean Normalised Distances (also bias invariant), ZND_P, like the Zero mean Normalised Sum of Squared Differences, ZNSSD (ZND_2).

Locally scaled distances – The aim of LD_P measures is to obtain the same mean of grey levels on each window: each grey level in the right image is scaled by the ratio between the left and right means. The two known measures are the Locally

scaled Sum of Absolute Differences, LSAD (LD_1), and the Locally scaled Sum of Squared Differences, LSSD (LD_2).

Variances – Two kinds of measures can be distinguished: the Variance of differences (bias invariant), VD, and the Variance of absolute P-powered differences, VAD_P, which gives the Variance Of Absolute Differences, VOAD (VAD_1), and the Variance Of Squared Differences, VOSD (VAD_2).

Fourth-order statistics-based measure – High order statistics have been investigated, and, in particular, by using a fourth-order cumulant of the grey level differences, K_4, designed to be robust against Gaussian noise [19].

NAME	ABBREVIATION	DEFINITION	VARIATION
Distances	D_P	$\|\mathbf{f}_l-\mathbf{f}_r\|_P{}^P$	$[0; \mathrm{I_{max}}{}^P N_f]$
Zero mean Distances	ZD_P	$\mathrm{D}_P(\mathbf{f}_l-\overline{\mathbf{f}_l},\mathbf{f}_r-\overline{\mathbf{f}_r})$	$[0; \mathrm{I_{max}}{}^P N_f]$
Normalised Distances	ND_P	$\frac{\mathrm{D}_P(\mathbf{f}_l,\mathbf{f}_r)}{\sqrt{\|\mathbf{f}_l\|_P{}^P \|\mathbf{f}_r\|_P{}^P}}$	$[0; \mathrm{I_{max}}{}^P N_f]$
Zero mean Normalised Distances	ZND_P	$\mathrm{ND}_P(\mathbf{f}_l-\overline{\mathbf{f}_l},\mathbf{f}_r-\overline{\mathbf{f}_r})$	$[0; \mathrm{I_{max}}{}^P N_f]$
Locally scaled Distances [7]	LD_P	$\mathrm{D}_P(\mathbf{f}_l,(\overline{\mathbf{f}_l}/\overline{\mathbf{f}_r})\mathbf{f}_r)$	$[0; \mathrm{I_{max}}{}^P N_f]$
Variance of Differences [18]	VD	$\mathrm{var}(\mathbf{f}_l-\mathbf{f}_r)$	$[0; \mathrm{I_{max}}{}^2]$
Variance of Absolute P-powered Differences [18]	VAD_P	$\mathrm{var}(\|\mathbf{f}_l-\mathbf{f}_r\|^P)$	$[0; \mathrm{I_{max}}{}^{2P}]$
Fourth-order statistics-based measure [19]	K_4	$\mathrm{m}((\mathbf{f}_l-\mathbf{f}_r)^4) - 3\,\mathrm{m}((\mathbf{f}_l-\mathbf{f}_r)^2)^2$	$[0; \mathrm{I_{max}}{}^4]$

Table 3
CLASSICAL family.

5 Derivative-based measures

All the measures of the DERIVATIVE family (Table 4) are based on the grey level distribution. They employ the derivatives of the images at different orders and use

the operators of Sobel, Roberts, Kirsch or Pratt. The gradient vector at $\mathbf{p}_w^{i,j}$ in I_w is $\nabla I_w^{i,j}$. The norm and the orientation are denoted respectively $\|\nabla I_w^{i,j}\|$ and $\theta_w^{i,j}$.

Seitz measures – The idea [22] of SEO_P (SEitz Operator) is to estimate the dissimilarity of the gradient vector directions by calculating the L_P norm of the gradient direction differences. These measures are efficient in the case of impulsive noise whereas they are not with Gaussian noise [7] and they are gain and bias invariant. We denote these measures by SES_P (SEitz Sobel) and SEK_P (SEitz Kirsch), with $P = 1, 2$. In fact, SES_1 was introduced in [22] whereas SES_2, SEK_1 and SEK_2 are improved versions proposed in [7].

Nishihara correlation, Nack measure and Pratt correlation – *Nishihara measure* [23], NIS, is the cross correlation, equation (1), of binary Laplacian images (similarity measure). It is not efficient with impulsive noise and occlusions [7]. For *Nack measures,* a convolution with the Roberts operator is applied. For each pixel, the region of interest (ROI) is binarized to take into account only 15% of the ROI (this percentage is empirically chosen by the authors). It allows them to be robust against noises and untextured areas because it takes into account only the most significant part of the ROI. These similarity measures [24], NA_m, $m = 1, 2$, are not robust against Gaussian or impulsive noises, but, the larger the correlation window, the better the results are [7]. In fact, NA_1 was proposed in [24] whereas NA_2 is a modified version proposed in [7] (we have also modified this measure in order to avoid division by zero). The similarity measure [25, pp. 666–667], PRATT, is ZNCC, cf. Table 2, applied to binary Laplacian values contained in $\mathrm{R}_{LoG}(\mathbf{f}_w)$. *Nishihara measure* is very "flexible", i.e. it gives the best score for the corresponding pixels and even for non-corresponding pixels. Consequently, this measure can be robust to occlusions or impulsive noise but can also induce errors. With the following binary vectors (for respectively $\mathbf{f}_l$ and one possible candidate $\mathbf{f}_r$), with

$O \in \{Rob, LoG\}$:

(a) $\mathrm{R}_O(\mathbf{f}_l) = (0\,0\,1\,1\,0\,1\,0\,1\,1)^T$ and

(b) $\mathrm{R}_O(\mathbf{f}_r) = (\mathbf{1}\,0\,1\,1\,0\,1\,0\,1\,1)^T$,

we obtain:

(r_1) $\mathrm{NIS}(\mathbf{f}_l, \mathbf{f}_l) = \mathrm{NIS}(\mathbf{f}_l, \mathbf{f}_r) = 5$ whereas

(r_2) $\mathrm{NA}_1(\mathbf{f}_l, \mathbf{f}_l) = 1$ and $\mathrm{NA}_1(\mathbf{f}_l, \mathbf{f}_r) = 5/6$.

For NIS, the best score is obtained with two correspondent vectors and two different vectors. In this example, even if each 0 is replaced by 1 between $\mathrm{R}_O(\mathbf{f}_l)$ and $\mathrm{R}_O(\mathbf{f}_r)$, $\mathrm{NIS}(\mathbf{f}_l, \mathbf{f}_r) = 5$. On the contrary, with:

(c) $\mathrm{R}_O(\mathbf{f}_r^{'}) = (0\,0\,\mathbf{0}\,1\,0\,1\,0\,1\,1)^T$,

we obtain:

(r_1) $\mathrm{NIS}(\mathbf{f}_l, \mathbf{f}_l) = 5$ and $\mathrm{NIS}(\mathbf{f}_l, \mathbf{f}_r^{'}) = 4$ whereas

(r_2) $\mathrm{NA}_1(\mathbf{f}_l, \mathbf{f}_l) = \mathrm{NA}_1(\mathbf{f}_l, \mathbf{f}_r^{'}) = 1$.

With the modified version of NA_1, we obtain:

(r_3) $\mathrm{NA}_2(\mathbf{f}_l, \mathbf{f}_l) = 1$ and $\mathrm{NA}_2(\mathbf{f}_l, \mathbf{f}_r^{'}) = 1/2$.

In this example, even if each 1 is replaced by 0 between $\mathrm{R}_O(\mathbf{f}_l)$ and $\mathrm{R}_O(\mathbf{f}_r^{'})$ until there is at least one 1 in $\mathrm{R}_O(\mathbf{f}_r^{'})$, $\mathrm{NA}_1(\mathbf{f}_l, \mathbf{f}_r^{'}) = 1$. The measure NA_2 reduces the problem of NA_1 with the introduction of 0. In conclusion, the measures of Nack are less "flexible" than the measure of Nishihara but they are not robust against occlusions.

Orientation code matching correlation – For this similarity measure, the gradient

direction code of pixel $\mathbf{p}_w^{i,j}$ is estimated by [26]:

$$\mathrm{C}_{OCM}(\mathbf{p}_w^{i,j}) = \begin{cases} \lfloor \theta_w^{i,j}/\Delta_\theta \rfloor & \text{if } (\|\nabla I_w^{i,j}\| > T_\Gamma) \\ L & \text{otherwise,} \end{cases}$$

where $\lfloor x \rfloor$ is the integer part of x and $\mathrm{C}_{OCM} \in \{0, 1, ..., N_{OCM} = 2\pi/\Delta_\theta, L\}$. The authors have used $\Delta_\theta = \pi/8$, $L = 255$, $T_\Gamma = 10$ (when the image range is $[0; 255]$) and the Sobel operator. The authors used L (here 255) in order to detect low-contrasted regions and, after, in the computation of the similarity cost, to reduce their influence. When the distance between the orientation codes of two pixels is estimated, if this distance is over N_{OCM} (only in the case when one of the two pixels has the value L), a constant value ($N_{OCM}/2$) is assigned, see Table 4 for the definition of this distance.

Gradient vector field correlation – For the previous measures, the gradient direction is only introduced: this might introduce errors, especially with low norm gradient vectors whose direction is not reliable. Consequently, a dissimilarity measure, the Gradient field Correlation, GC, bias invariant measure, is introduced in [27].

6 Non-parametric measures

This family (Table 5) is based on non-parametric transformations, i.e. no hypothesis about the grey level distributions is made. We distinguish χ^2 and Jeffrey measures [3], Kaneko measures [28,29], Zabih measures [30] and ordinal measures [31].

χ^2 and Jeffrey measures – These dissimilarity measures [3] are used for segmentation and they seem to be interesting for correlation-based matching. For image

NAME	ABBREVIATION	DEFINITION	VARIATION
Seitz [22,7]	SEO_P	$\|\mathrm{R}_O(\mathbf{f}_l)-\mathrm{R}_O(\mathbf{f}_r)\|_P{}^P$	$[0; \mathrm{I}_{\max}{}^P N_f]$
Nishihara [23]	NIS	$\mathrm{R}_{LoG}(\mathbf{f}_l)\cdot\mathrm{R}_{LoG}(\mathbf{f}_r)$	$[0; N_f]$
Nack 1 [24]	NA_1	$\frac{\mathrm{R}_{Rob}(\mathbf{f}_l)\cdot\mathrm{R}_{Rob}(\mathbf{f}_r)}{N_f\ \mathrm{m}(\mathrm{R}_{Rob}(\mathbf{f}_r))}$	$[0; 1]$
Nack 2 [7]	NA_2	$\frac{\mathrm{NA}_1(\mathbf{f}_l,\mathbf{f}_r)}{N_f\,\mathrm{m}(\mathrm{R}_{Rob}(\mathbf{f}_l))-\mathrm{R}_{Rob}(\mathbf{f}_l)\cdot\mathrm{R}_{Rob}(\mathbf{f}_r)+1}$	$[0; 1]$
Pratt [25]	PRATT	$\mathrm{ZNCC}(\mathrm{R}_{LoG}(\mathbf{f}_l),\mathrm{R}_{LoG}(\mathbf{f}_r))$	$[-1; 1]$
Orientation Code Matching [26]	OCM	$\frac{1}{N_f}\mathrm{D}_{OCM}(\mathrm{R}_{OCM}(\mathbf{f}_l),\mathrm{R}_{OCM}(\mathbf{f}_r)$	$\left[0; \frac{N_{OCM}}{2}\right]$
Gradient vector field Correlation [27]	GC	$\frac{\sum_{p=-N_v}^{N_v}\sum_{q=-N_h}^{N_h}\|\nabla I_l^{i+p,j+q}-\nabla I_r^{v+p,w+q}\|}{\sum_{p=-N_v}^{N_v}\sum_{q=-N_h}^{N_h}(\|\nabla I_l^{i+p,j+q}\|+\|\nabla I_r^{v+p,w+q}\|)}$	$[0; \infty[$

Table 4
DERIVATIVE family – The vectors $\mathrm{R}_O(\mathbf{f}_w)$ contain the gradient directions of $\mathbf{f}_w$ after using the Sobel or Kirsch operator, $\mathrm{R}_{LoG}(\mathbf{f}_w)$ is the correlation window in the binary Laplacian images, $\mathrm{R}_{Rob}(\mathbf{f}_w)$ contains the binary values after the Roberts transformation and $\mathrm{R}_{OCM}(\mathbf{f}_w) = (\cdots \mathrm{C}_{OCM}(f_w^k)\cdots)^T$ are the orientation codes in $\mathbf{f}_w$, cf. section 5. As the OCM code is cyclic, the maximal distance is $N_{OCM}/2$ and the distance $\mathrm{D}_{OCM}(\mathbf{f}_l, \mathbf{f}_r)$ is $\sum_{k=0}^{N_f-1}\mathrm{D}(f_l^k, f_r^k)$, with $\mathrm{D}(a,b) = \min\{|a-b|,\ N_{OCM}-|a-b|\}$ if $|a-b| < N_{OCM}$ $N_{OCM}/2$ otherwise.

retrieval, they are numerically stable, symmetric and robust with respect to noise.

Increment Sign Correlation – The similarity measure, ISC [28], gain and bias invariant, uses: $\mathbf{b}_w = \left(\ldots b_w^k \ldots\right)^T$ with $k = 0\cdot\cdot N_f - 1$ and

$$b_w^k = \begin{cases} 1 & \text{if } (k < (N_f - 1)) \text{ and } (f_w^{k+1} \geq f_w^k) \\ 0 & \text{otherwise.} \end{cases}$$

If the grey level increases between f_w^k and f_w^{k+1}, then b_w^k equals 1. The vectors $\mathbf{b}_l$ and $\mathbf{b}_r$ are compared to estimate how the variations of the grey levels are similar.

Selective Correlation Coefficient, SCC – This measure is a variant of ISC and has the same properties [29]. It is based on these weights (determined with $\mathbf{b}_w$):

$$\mathbf{e} = (\cdots e^k \cdots)^T,\ k = 0 \cdot\cdot N_f - 1,\ e^k = \begin{cases} 1 - |b_l^k - b_r^k| & \text{if } k = 0 \text{ or } k \text{ even} \\ e^{k-1} & \text{otherwise.} \end{cases}$$

If grey levels change in the same direction between f_l^k and f_l^{k+1}, and between f_r^k and f_r^{k+1} then e_l^k and e_l^{k+1} (k is even) equal 1.

Rank measures – The rank transformation, rk, is the number of pixels in $\mathbf{f}_w$ with a grey level lower than the grey level of the central pixel of $\mathbf{f}_w$: $rk(\mathbf{p}_w^{i,j}) = \mathrm{card}(\{I_w^{i+p,j+q} \mid I_w^{i+p,j+q} < I_w^{i,j}\})$. In consequence, the window transform $\mathrm{R}_{rk}(\mathbf{f}_w)$ is defined by:

$\mathrm{R}_{rk}(\mathbf{f}_w) = (\ldots\ rk(\mathbf{p}_w^{i+p,j+q})\ \ldots)$, $p \in [-N_v; N_v],\ q \in [-N_h; N_h]$ and

An illustration of the estimation of $rk(\mathbf{p}_w^{i,j})$ and $\mathrm{R}_{rk}(\mathbf{f}_w)$ is given in Figure 1. Zabih and Woodfill [30] employ the L_1 and L_2 norms (RANK_P in Table 5). The measures are gain and bias invariant.

100	30	20	10	10
120	*10*	*25*	*15*	10
130	*40*	**43**	*46*	40
120	*49*	*70*	*50*	40
140	40	60	40	40

Fig. 1. Illustration for estimating the Rank transform – If the Rank transform is estimated for the pixel **43** and its 3×3 neighborhood (in italic), we obtain: $rk(\mathbf{p}_w^{i,j}) = 4$ for this pixel and $\mathrm{R}_{rk}(\mathbf{f}_w) = (0\ 4\ 3\ 2\ \mathbf{4}\ 6\ 3\ 8\ 6)^T$ for his neighborhood.

Census measure – This similarity measure, proposed in [30], uses a transform that produces a bit chain which represents the pixels with an intensity lower than the central pixel: $\mathrm{R}_\tau(\mathbf{f}_w) = \bigotimes_{k \in [0; N_f - 1]} \xi(f_w^{N_f/2}, f_w^k)$ where $\xi(f_w^{N_f/2}, f_w^k) = 1$ if $f_w^k < f_w^{N_f/2}$. CENSUS is the sum of the Hamming distances between the codes of each pixel of the correlation window. It is gain and bias invariant.

Ordinal measures – A similarity measure model α, which is gain and bias invariant, is defined by: $\alpha(\mathbf{f}_l, \mathbf{f}_r) = 1 - (2\mathrm{D}_\alpha(\mathrm{R}_\pi(\mathbf{f}_l), \mathrm{R}_\pi(\mathbf{f}_r)))/\mathrm{D}_{\max}$, where $\mathrm{R}_\pi(\mathbf{f}_w)$ contains the ranks of the pixels in $\mathbf{f}_w$, D_α is a distance and $\mathrm{D}_{\max}$ is the maximum of D_α. In [31], they tested the Hamming distance, the Kendall and the Spearman measures that are not effective and they proposed the following distances.

The ranks of the element of $\mathbf{f}_w$ are stored in $\mathrm{R}_\pi(\mathbf{f}_w)$, a permutation of $(1\ 2\ \ldots\ N_f)^T$. The element i of a composition of permutations is given by: $\mathrm{Comp}^i = \mathrm{R}_\pi^k(\mathbf{f}_r)$ with $k = \mathrm{R}_\pi^i(\mathbf{f}_l)^{-1}$ and $\mathrm{R}_\pi(\mathbf{f}_l)^{-1}$ is the inverse permutation of $\mathrm{R}_\pi(\mathbf{f}_l)$. With the example of Figure 1, the ranks are:

$\mathrm{R}_\pi(\mathbf{f}_l) = (1\ 3\ 2\ 4\ 5\ 6\ 7\ 9\ 8)^T$, and if we study this candidate:

$\mathbf{f}_r = (55\ 20\ 21\ 40\ 18\ 46\ 49\ 15\ 50)^T$ with

$\mathrm{R}_\pi(\mathbf{f}_r) = (9\ 3\ 4\ 5\ 2\ 6\ 7\ 1\ 8)^T$, then

$\mathrm{Comp}(\mathrm{R}_\pi(\mathbf{f}_l), \mathrm{R}_\pi(\mathbf{f}_r)) = (9\ 4\ 3\ 5\ 2\ 6\ 7\ 8\ 1)^T$.

The element i of the deviation is:

$$\mathrm{Dev}^i = \sum_{j=0}^{j=i} J(\mathrm{Comp}^j > (i+1)) \text{ with } J(B) = \begin{cases} 1 & \text{if B is true or} \\ 0 & \text{otherwise.} \end{cases}$$

With the previous example, we have:

$\mathrm{Dev}(\mathrm{R}_\pi(\mathbf{f}_l), \mathrm{R}_\pi(\mathbf{f}_r)) = (1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 0)^T$. The κ measure is based on the maximum in $\mathrm{Dev} = (\ldots \mathrm{Dev}^k \ldots)^T$:

$$\kappa(\mathbf{f}_l, \mathbf{f}_r) = 1 - (2 \max_{k=0..N_f-1} \mathrm{Dev}^k)/\lfloor N_f/2 \rfloor.$$

A variant, less expensive than the κ measure, is the χ measure:

$$\chi(\mathbf{f}_l, \mathbf{f}_r) = 1 - (2\text{Dev}^{N_f/2})/\lfloor N_f/2 \rfloor.$$

The ordinal measures are invariant to gain and bias and tolerate factionalism, i.e. they are robust against outliers and, so, against occlusions. However, they are "flexible", like the derivative-based measures, and they can produce errors in areas without occlusion. In fact, a maximal correlation score can be obtained even if the two correlation windows are not strictly identical. For example, with $\mathbf{f}_l = (0\ 1\ 22\ 35\ 46\ 58\ 61\ 121\ 123)^T$ and $\mathbf{f}_r = (0\ 2\ 42\ 60\ 81\ 100\ 123\ 124\ 125)^T$, we have $\kappa(\mathbf{f}_l, \mathbf{f}_l) = \kappa(\mathbf{f}_l, \mathbf{f}_r) = 1$. The maximal score is reached with $(\mathbf{f}_l, \mathbf{f}_l)$ but also with $(\mathbf{f}_l, \mathbf{f}_r)$. Consequently, erroneous correspondences can be obtained.

NAME	ABBREVIATION	DEFINITION	VARIATION
χ^2 measure [3]	χ^2	$\sum_{k=0}^{N_f-1} \frac{2(f_l^k - f_r^k)^2}{f_l^k + f_r^k}$	$[0; \mathrm{I}_{\max} N_f]$
Jeffrey measure [3]	JEFF	$\sum_{k=0}^{N_f-1} f_l^k \log(\frac{2f_l^k}{f_l^k + f_r^k}) + f_r^k \log(\frac{2f_r^k}{f_l^k + f_r^k})$	$[0; \mathrm{I}_{\max} N_f]$
Increment Sign Correlation [28]	ISC	$\frac{1}{N_f - 1}(\mathbf{b}_l \cdot \mathbf{b}_r + (\mathbf{1} - \mathbf{b}_l) \cdot (\mathbf{1} - \mathbf{b}_r))$	$[0; 1]$
Selective Coefficient Correlation [29]	SCC	$\frac{(\mathbf{E}(\mathbf{f}_l - \overline{\mathbf{f}_l}) \cdot (\mathbf{f}_r - \overline{\mathbf{f}_r}))}{\|\mathbf{E}(\mathbf{f}_l - \overline{\mathbf{f}_l})\| \|\mathbf{E}(\mathbf{f}_r - \overline{\mathbf{f}_r})\|}$	$[0; 1]$
Rank [30]	RANK_P	$\|\mathrm{R}_{rk}(\mathbf{f}_l) - \mathrm{R}_{rk}(\mathbf{f}_r)\|_P{}^P$	$[0; N_f^{P+1}]$
Census [30]	CENSUS	$\sum_{k=0}^{N_f-1} \mathrm{D}_H(\mathrm{R}_\tau(\mathbf{f}_l), \mathrm{R}_\tau(\mathbf{f}_r))$	$[0; N_f]$
Ordinal measures [31]	α	$1 - 2\frac{\mathrm{D}_\alpha(\mathrm{R}_\pi(\mathbf{f}_l), \mathrm{R}_\pi(\mathbf{f}_r))}{\mathrm{D}_{\max}}$	$[-1; 1]$

Table 5
NON-PARAMETRIC family – The Hamming distance is: $\mathrm{D}_H(\mathbf{f}_l, \mathbf{f}_r) = \sum_{i=0}^{N_f-1} \mathrm{sgn}\,|f_l^i - f_r^i|$, with $\mathrm{sgn}(x) = 1$ if $x > 0$, 0 if $x = 0$ or -1 otherwise. The diagonal matrix $\mathbf{E}$ contains the values e^k, with $k \in [0; N_f - 1]$.

7 Robust measures

We are particularly concerned with the occlusion problem which appears in the vicinity of a pixel near a depth discontinuity. In fact, some pixels lie on a first level of depth whereas the other pixels lie on a second level. It can disturb the matching process and introduce erroneous matches. To take this problem into account, the measures of the ROBUST family, cf. Tables 6 and 9, consider pixels with a depth different to the main pixel as outliers (Figure 2). So, they employ the tools of robust statistics that are less sensitive to outliers than classical ones.

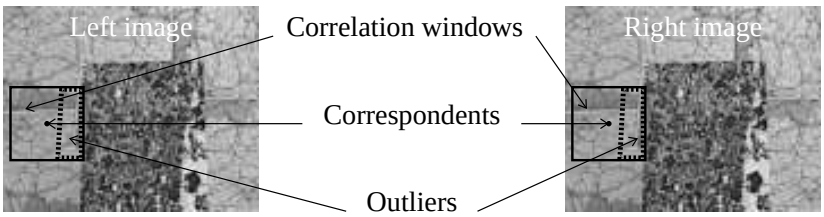


Fig. 2. Robust measure principle – The small disks are the two correspondent pixels. The squares are the correlation windows and the dashed polygons are the parts of the correlation windows which contain very different grey levels. In fact, in the left image, this part belongs to the background whereas, in the right, it belongs to the foreground. Therefore, comparing these two areas is not relevant.

Partial correlation – The principle of the similarity partial correlation is to calculate a score with different weights for each pixel pair. These weights are chosen according to the degree of membership of the set of correct correspondent pairs: the higher the degree, the greater the weight. The matrices $\mathbf{B}_w$ contain the weights β_w^k, with $k \in [0; N_f - 1]$, applied to $\mathbf{f}_w$ and this function is defined: $\mathrm{R}_m(\mathbf{f}_w) = \mathbf{B}_w \cdot \mathbf{f}_w$ where m corresponds to the chosen method for calculating the weights. The measures based on partial correlation are noted RM for Reweighted Measures and, depending on the used measure, RM might be gain and bias invariant. Lan [9] supposes that, for two correspondent pixels, the grey levels of the two correlation windows, without occlusion, are linearly dependent. It allows one to take into account the intensity distortions and to include a Gaussian white noise. In other words, if

the grey levels of two pixels in the same position in the two windows are considered as the coordinates in a plane, $\mathbf{a}^k = (f_l^k \; f_r^k)^T$, then a set of aligned points is obtained and this cloud of points is only disturbed by "normal" noise. If there are occluded pixels, it is assumed that the grey levels in the same position are very different and, so, these pairs do not respect the previous model: they are outliers. For that reason, a line fitting technique can be employed to detect outliers and two robust estimators have been proposed: LMS, Least Median of Squares for the Reweigthed Zero mean Sum of Squared Differences correlation, RZSSD, and MVE, Minimum Volume Ellipsoid for the Reweighted Zero mean Normalised Cross Correlation, RZNCC.

A random sampling with Monte-Carlo method is needed. Some robust parameter estimators are based on a minimisation with no explicit solution, like LMS and MVE. A first idea is to build minimal subsets of data (it contains the lowest number of data that is needed to calculate the parameters), then to estimate the parameters for each subsets and finally to select those which minimise the global criterion. The major drawback is the number of subsets q: with n data and subsets of m data, we have C_m^n combinations. Another possibility is to randomly choose a limited number of subsets. It depends on the risk of not finding the global minimum, more precisely: $\mathrm{Prob} = 1 - (1 - (1 - \epsilon)^m)^q$, ϵ is the maximal proportion of outliers in the initial data and Prob is the probability that, at least, one subset is correct, i.e. with no outliers. By choosing a priori $\epsilon = 0.5$ and $\mathrm{Prob} = 0.95$, q is defined by:

$$q = \frac{\log(1 - \mathrm{Prob})}{\log\left(1 - (1 - \epsilon)^m\right)}.$$

Consequently, we have to randomly select 11 subsets for LMS ($m = 2$) and 23 for MVE ($m = 3$).

The aim of LMS estimator is to find the parameters of the line which minimise: $\mathrm{med}_{k=0 \cdot \cdot N_f - 1}(r^k)^2$ where r^k is the Euclidean distance between the line and $\mathbf{a}^k$. The

weights are obtained by thresholding the distance between the point and the line where the threshold is a robust estimation of the standard deviation:

$$w_{LMS}^k = \begin{cases} 1 & \text{if } |r^k|/\widehat{\sigma} \leq 2.5 \\ 0 & \text{otherwise} \end{cases} \text{ with } \widehat{\sigma} = 1.4826 \left(1 + \frac{5}{N_f - 2}\right) \sqrt{\underset{k=0\cdot\cdot N_f-1}{\text{med}} (r^k)^2}.$$

The factor 1.4826 permits an estimation without bias with Gaussian noise and the term $1 + 5/(N_f - 2)$ allows correction with small subsets [32].

For the second measure, the estimation of the MVE of N_f points, $\mathbf{a}^k = (f_l^k \ f_r^k)^T$, is introduced in [9]. The authors estimate the parameters of the MVE with h points (here, $h = \lfloor N_f/2 \rfloor + 1$) that are represented by the coordinates of the centre $\overline{\mathbf{a}}$ of the ellipsoid and the covariance matrix $\mathbf{A}$. The parameters have to minimize

$\sqrt{\det(\mathbf{A})}$ with $\mathbf{A} = (\chi^2_{2;0.5})^{-1} m^2 \mathbf{A}'$ where

$m^2 = \underset{k=0\cdot\cdot N_f-1}{\text{med}} (\mathbf{a}^k - \overline{\mathbf{a}})^T \mathbf{A}'^{-1} (\mathbf{a}^k - \overline{\mathbf{a}})$. Using χ^2 allows to produce an estimation being robust against Gaussian noise. By introducing Mahalanobis distance, weights are:

$$w_{MVE}^k = \begin{cases} 1 & \text{if } (\mathbf{a}^k - \widehat{\overline{\mathbf{a}}})^T \widehat{\mathbf{A}}^{-1} (\mathbf{a}^k - \widehat{\overline{\mathbf{a}}}) \leq \chi^2_{2;0.975} \\ 0 & \text{otherwise,} \end{cases}$$ where the $\chi^2_{2;0.975}$ is a commonly used threshold for detecting outliers, and, $\widehat{\overline{\mathbf{a}}}$ and $\widehat{\mathbf{A}}$ are the estimated parameters.

Partial correlations, RZSSD and RZNCC, are robust against occlusions [9].

Robust ZNCC measures – Some measures are a robust version of ZNCC, cf. Table 6, like the quadrant correlation, QUAD [33, pp. 204–205], and the measure of Trujillo [34]. They are gain and bias invariant also. The quadrant transformation is applied on the vectors $\mathbf{f}_w$ and gives binary values. For the Trujillo, the mean (used

for centering) is replaced by a median and the normalisation with L_2 becomes L_1.

Pseudo-norms – These dissimilarity measures are robust distances [35], i.e. L_P norms with $0 < P < 1$. The adhesion effect occurs at discontinuity boundaries induced by an occlusion and the consequence is the dilation of the occluding object in the disparity map [2] . It appears with classical norms while the pseudo-norms alleviate it because, with $P > 1$, the greater the power, the more important influence of the pixels that induce large grey level differences. With a pseudo-norm, the lower the power (near 0), the less important influence the differences, i.e. they are robust against occlusions. Unfortunately, pseudo-norms generate the erosion of object corners.

NAME	ABBREVIATION	DEFINITION	VARIATION
Partial correlation [9]	RM_m	$\text{Mes}(\text{R}_m(\mathbf{f}_l),\text{R}_m(\mathbf{f}_r))$	$[-1;1]$
Quadrant correlation [33]	QUAD	$\text{ZNCC}(\text{R}_{quad}(\mathbf{f}_l),\text{R}_{quad}(\mathbf{f}_r))$	$[0;1]$
Robust ZNCC [34]	ZNCC_R	$\frac{(\mathbf{f}_l-\text{med}(\mathbf{f}_l))\cdot(\mathbf{f}_r-\text{med}(\mathbf{f}_r))}{\|\mathbf{f}_l-\text{med}(\mathbf{f}_l)\|_1\|\mathbf{f}_r-\text{med}(\mathbf{f}_r)\|_1}$	$[-1;1]$
Pseudo-norm [35]	D_P	$\|\mathbf{f}_l-\mathbf{f}_r\|_P{}^P$ with $0<P<1$	$[0;+\infty[$

Table 6
ROBUST family (state of the art) – For the quadrant correlation, the authors suggest to use this transform: $\text{R}_{quad}(\mathbf{f}_l) = \text{sgn}\left(\frac{(\mathbf{f}_l-\text{med}(\mathbf{f}_l)}{\text{med}\,|\mathbf{f}_l-\text{med}(\mathbf{f}_l)|}\right)$. In fact, the divisor is anyway positive and does not affect the sign, and, $\text{R}_{quad}(\mathbf{f}_l) = \text{sgn}(\mathbf{f}_l - \text{med}(\mathbf{f}_l))$ is more appropriate.

8 Proposed robust measures

We propose to complete the set of existing robust measures using again the principle illustrated in Figure 2.

Robust variance, MAD – This dissimilarity measure, Median Absolute Deviation, is a robust estimation of the variance of the grey level differences. We can consider

[2] Disparity is the displacement between a pixel in one of the image and its correspondent in the other image.

it as a robust version of ZNSSD and it is gain and bias invariant.

Least Median of Powers (LMP) – We proposed the LMP_P, a generalisation to P powers, of the least median of squares [32] . It is a robust version of D_P and an alternative to the pseudo-norms, cf. § 7. It is gain and bias invariant.

Least Trimmed Powers (LTP) – It is based on the least trimmed squares [32] where the squared grey level differences are sorted and the h first values (here, $h = N_f/2$) are summed. Instead of using the squared difference, LTP can be defined with any power difference. It is gain and bias invariant.

Smooth Median Powered Deviation (SMPD) – The Smooth Median Absolute Deviation, SMAD [36], is also a robust estimation of the variance. The measure SMPD_P is a normalised version of LTP_P with bias invariance only.

M-estimators – The least mean of squares estimation is sensitive to outliers, while the M-estimators use a criterion which replaces the square by an object function, ρ_m, symmetric with a single minimum at 0 [37], less sensitive to outliers because it increases less quickly than the square function. So, we suggest ρ-based dissimilarity measures, ME_m, cf. Table 7 and Figure 3, which are bias invariant.

R-estimators – The principle is to change the square function by weights, given by the J_m functions, that depend on the rank of the differences. This principle helps to decrease the influence of the outliers and we propose dissimilarity measures based on R-estimators, RE_m, cf. Table 8 and Figure 3 [32,38,39]. The integral of J_m on its definition domain must be equal to 0.These measures are gain and bias invariant.

NAME	FUNCTION	NAME	FUNCTION
$L_1 - L_2$	$\rho_1(x)=(\sqrt{1+x^2}-1)/2$	Fair	$\rho_2(x)=\lvert x\rvert-\log(1+\lvert x\rvert)$
Cauchy	$\rho_3(x)=\log(1+x^2)$	Geman-McClure	$\rho_4(x)=\frac{x^2}{2(1+x^2)}$

NAME	FUNCTION
Welsh	$\rho_5(x)=(1-\mathrm{e}^{-x^2})$
Tukey	$\rho_6(x)=\begin{cases}(1-(1-x^2)^6) & \text{if } \lvert x\rvert\leq 1\\ 1 & \text{otherwise}\end{cases}$
Huber	$\rho_7(x)=\begin{cases}(x^2)/2 & \text{if } \lvert x\rvert\leq 1.345\\ 1.345(\lvert x\rvert-1.345/2) & \text{otherwise}\end{cases}$
Rousseeuw	$\rho_8(x)=2\log(e^x+1)-x-2\log(2)$

Table 7
The ρ_m functions of the M-estimators measures.

NAME	FUNCTION	NAME	FUNCTION
Wilcoxon	$J_1(t)=t-1/2$	Median	$J_2(t)=\mathrm{sgn}(t-1/2)$

NAME	FUNCTION
Van der Waerden	$J_3(t)=\phi^{-1}(t)$
Optimal B-robust estimator	$J_4(t)=\begin{cases}-1.4634 & \text{if } 0\leq t\leq 0.39\\ 1.47\phi^{-1}(t) & \text{if } 0.39<t\leq 0.61\\ 1.4634 & \text{if } 0.61<t\leq 1\end{cases}$
Minimax	$J_5(t)=\begin{cases}-1.14 & \text{if } 0\leq t\leq 0.48\\ \phi^{-1}(0.5+\frac{t-0.5}{t-0.1}) & \text{if } 0.48<t\leq 0.52\\ 1.14 & \text{if } 0.52<t\leq 1\end{cases}$

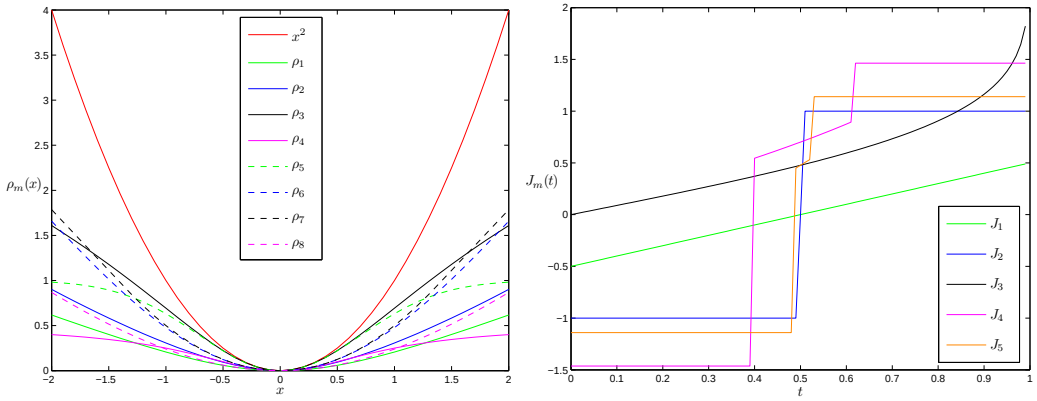
Table 8
The J_m functions of the R-estimator measures – The ϕ function is the normal distribution function. The values of ϕ^{-1} lie on $[\phi^{-1}_{\text{min}};\phi^{-1}_{\text{max}}]$ and $t\in[0;1]$.

9 Evaluation and comparison protocol

We propose a protocol designed specifically for comparing correlation-based methods and for describing their behaviour with occlusions, and, we present the tested images, the evaluation areas, the criteria and how all the results are summarized.

NAME	ABBREVIATION	DEFINITION	VARIATION
Robust variance [1]	MAD	$\text{med}\,\lvert(\mathbf{f}_l-\mathbf{f}_r)-\text{med}(\mathbf{f}_l-\mathbf{f}_r)\rvert$	$[0;+\infty[$
Least Median of Powers [1]	LMP_P	$\text{med}\ (\lvert\mathbf{f}_l-\mathbf{f}_r\rvert^P)$	$[0;\text{I}_{\text{max}}{}^P]$
Least Trimmed Powers [1]	LTP_P	$\sum_{k=0}^{h-1}(\lvert\mathbf{f}_l-\mathbf{f}_r\rvert^P)_{k:N_f-1}$	$[0;\text{I}_{\text{max}}{}^P h]$
Smooth Median Powered Deviation [1]	SMPD_P	$\sum_{k=0}^{h-1}(\lvert\mathbf{f}_l-\mathbf{f}_r-\text{med}(\mathbf{f}_l-\mathbf{f}_r)\rvert^P)_{k:N_f-1}$	$[0;\text{I}_{\text{max}}{}^P h]$
M-estimator [1]	ME_m	$\sum_{k=0}^{N_f-1}\rho_m(f_l^k-f_r^k)$	$[\rho_m^{\text{min}}N_f;$ $\rho_m^{\text{max}}N_f]$
R-estimator [1]	RE_m	$\sum_{k=0}^{N_f-1}J_m\left(\frac{\text{R}_\pi(f_l^k-f_r^k)}{N_f-1}\right)(f_l^k-f_r^k)$	$[\text{J}_m^{\text{min}}\text{I}_{\text{max}};$ $\text{J}_m^{\text{max}}\text{I}_{\text{max}}]$

Table 9
ROBUST family (proposed measures) – The ordered values of $\mathbf{f}_w$ are represented by: $(f_w)_{0:N_f-1} \leq \ldots \leq (f_w)_{N_f-1:N_f-1}$. The notation $|\mathbf{f}_l - \mathbf{f}_r|^P$ means $(\ldots|\mathbf{f}_l^k - \mathbf{f}_r^k|^P \ldots)^T$. The terms ρ_m^{min} and ρ_m^{max} are the lower and the upper bounds of ρ_m and J_m^{min} and J_m^{max} are the lower and the upper bounds of J_m on $[0;\text{I}_{\text{max}}]$. The rank of $(f_l^k - f_r^k)$ is stored in $\text{R}_\pi(f_l^k - f_r^k)$.



(a) (b)

Fig. 3. Visualisation of the functions used by M-estimators (a) and R-estimators (b) – It highlights how the influence of the highest grey level differences is reduced by using these functions compared to square function (for M-estimator, in (a), this function has been plotted).

Tested images – Forty two images are tested (examples are given in Figure 5): a random-dot stereogram (number 1), two synthetic pairs (numbers 2 and 3), one real pair made by Bocquillon [3] (number 4) and thirty eight real pairs introduced by

[3] http://www.irit.fr/ Benoit.Bocquillon/MYCVR/research.php

Scharstein and Szeliski: six have been proposed in 2002 [40] (number 7 to 12), two in 2003 [41] (numbers 13 and 14), six in 2005 [42] (numbers 15 to 20) and twenty two in 2006 [43] (numbers 21 to 42). The last ones, the set of 2005 and 2006, are the most complex images. Compared to the protocol of Scharstein and Szeliski, our protocol has the advantage of presenting the results obtained for forty-two images and not only for four images. In section 10, for reason of space, we only present a sample of the obtained results but all the tests have been done on all the cited images. Moreover, to test robustness against Gaussian noise and impulsive noise, auto-correlation has been performed for two kinds of images: a *random* image and a real image, *Sand*, of size 128×128, with 5% and 10% of impulsive noise or a Signal to Noise Ratio (SNR) equal to 0.1 and 0.2 for Gaussian noise, cf. Figure 4.

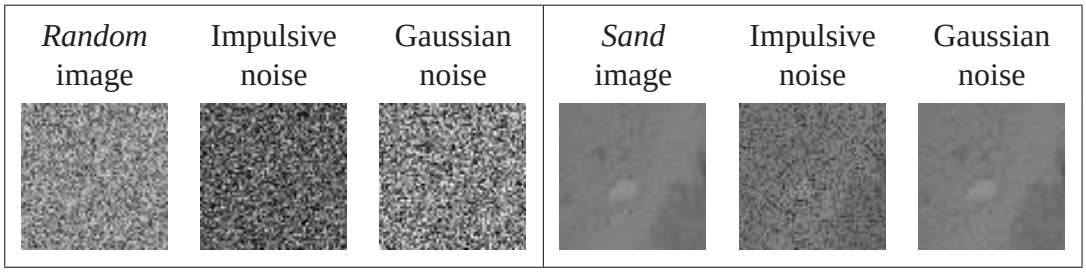


Fig. 4. Tested images with different type of noises.

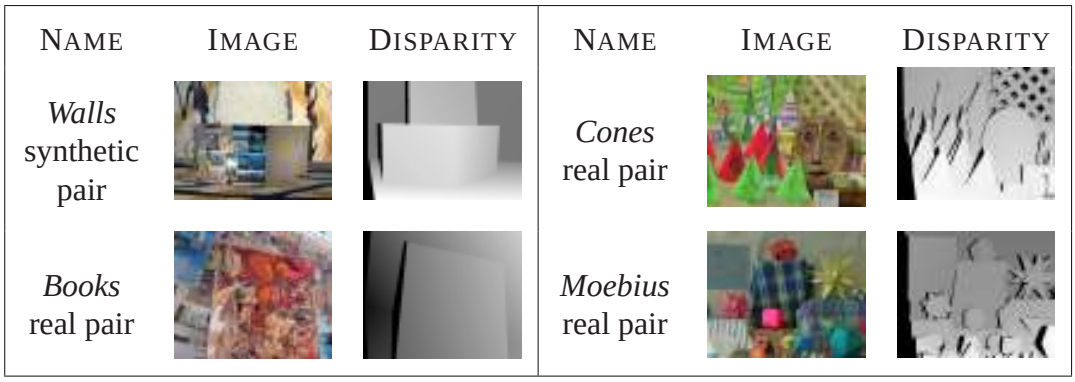


Fig. 5. Examples of data used in our tests (left images and disparities).

Evaluation areas – The advantage of the protocol is to quantify the behaviour of the methods based on correlation measure near occlusions and discontinuities, this is why we consider these areas, see Figure 6:

- *Occluded pixels* – They are pixels without correspondent:

$$\mathrm{O}(\mathbf{p}_w^{i,j}) = \begin{cases} 1 & \text{if } \mathbf{p}_w^{i,j} \text{ is an occluded pixel} \\ 0 & \text{otherwise.} \end{cases}$$

- *Occlusion area* – $\mathrm{OA}(I_w)$ contains all the occluded pixels in I_w:

$$\mathrm{OA}(I_w) = \left\{ \mathbf{p}_w^{i,j} \mid \mathrm{O}(\mathbf{p}_w^{i,j}) = 1 \right\}.$$

- *Pixels near occluded pixels* – They are the pixels in the neighbourhood of occluded pixels. This vicinity is related to the size of the correlation window: it corresponds to the morphological dilation of the occlusion area using the correlation window as structuring element:

$$\mathrm{NO}(\mathbf{p}_w^{i,j}) = \begin{cases} 1 & \text{if } (\mathrm{O}(\mathbf{p}_w^{i,j}) = 0) \text{ and } (\mathrm{V}(\mathbf{p}_w^{i,j}) = \sum_{\mathbf{p}_w^{i',j'} \in \mathrm{W}(\mathbf{p}_w^{i,j})} \mathrm{O}(\mathbf{p}_w^{i',j'})) > 0) \\ 0 & \text{otherwise,} \end{cases}$$

  where $\mathrm{W}(\mathbf{p}_w^{i,j})$ is the set that contains all the pixels of the correlation window.

- *Occlusion influence area* – $\mathrm{OIA}(I_w)$ contains all the pixels near occluded pixels in I_w: $\mathrm{OIA}(I_w) = \{\mathbf{p}_w^{i,j} \mid \mathrm{NO}(\mathbf{p}_w^{i,j}) = 1\}$.
- *Whole occlusion area* – $\mathrm{WOA}(I_w)$ is the union of OA and OIA for I_w: $\mathrm{WOA}(I_w) = \mathrm{OA}(I_w) \cup \mathrm{OIA}(I_w) = \{\mathbf{p}_w^{i,j} \mid (\mathrm{O}(\mathbf{p}_w^{i,j}) = 1) \text{ or } (\mathrm{NO}(\mathbf{p}_w^{i,j}) = 1)\}$.
- *Pixels near discontinuity* – They are in the vicinity of a discontinuity:

$$\mathrm{ND}(\mathbf{p}_w^{i,j}) = \begin{cases} 1 & \text{if } \exists\, \mathbf{p}_w^{i',j'} \in \mathrm{W}(\mathbf{p}_w^{i,j}) \;\Big|\; \|d(\mathbf{p}_w^{i',j'}) - d(\mathbf{p}_w^{i,j})\| \geq 1 \\ 0 & \text{otherwise,} \end{cases}$$

  where $d(\mathbf{p}_w^{i,j})$ is the disparity of $\mathbf{p}_w^{i,j}$. We select all the pixels that induce a differ-

ence of disparity in its neighborhood (this is why the threshold equals 1).

- *Discontinuity area* – It corresponds to: $\mathrm{DA}(I_w) = \{\mathbf{p}_w^{i,j} \mid \mathrm{ND}(\mathbf{p}_w^{i,j}) = 1\}$.

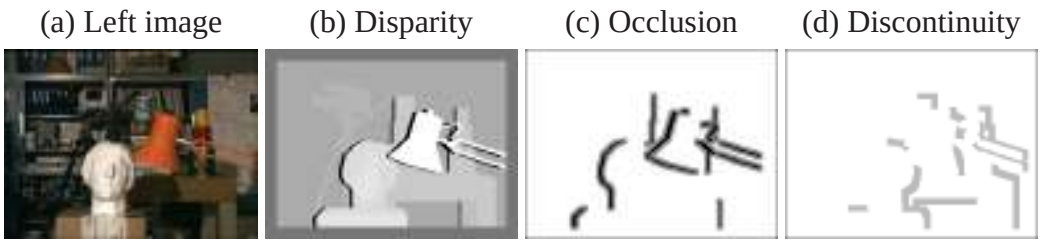


Fig. 6. Occlusion areas – We have calculated the Whole Occlusion Area WOA (c), with the ground truth (b), a disparity map (each pixel represents the disparity range defined by the distance between the position of the pixel in the left image and its correspondent in the right image. The brighter the pixel, the closer the point to the image plane and the larger the disparity). In (c), grey pixels are in the Occlusion Influence Area, OIA, and black pixels are in the occlusion area, OA. In (d), grey pixels are pixels near discontinuities.

Evaluation criteria – The theoretical disparity function is d_{ref} and the error is $\mathrm{Err}_w^{i,j} = \|d(\mathbf{p}_w^{i,j}) - d_{\text{ref}}(\mathbf{p}_w^{i,j})\|$, significant only if $d(\mathbf{p}_w^{i,j}) \neq occ$ and $d_{\text{ref}}(\mathbf{p}_w^{i,j}) \neq occ$, where occ is the value for occluded pixels. We calculate the following percentages:

(1) *Correct matches*, COR – A match is correct if $\mathrm{Err}_w^{i,j} < 1$.

(2) *Accepted matches*, ACC – A match is accepted if $1 \leq \mathrm{Err}_w^{i,j} < 2$.

(3) *Bad matches*, BAD – A match is bad if $2 \leq \mathrm{Err}_w^{i,j} < 3$.

(4) *Erroneous matches*, ERR – A match is erroneous if $\mathrm{Err}_w^{i,j} \geq 3$.

(5) *False positive, false negative,* FPO and FNE – The method estimates that the pixel is matched whereas it is not matched and vice versa. A match is a false positive (respectively a false negative) if $d(\mathbf{p}_w^{i,j}) \neq d_{\text{ref}}(\mathbf{p}_w^{i,j})$ and $d_{\text{ref}}(\mathbf{p}_w^{i,j}) = occ$ (respectively $d(\mathbf{p}_w^{i,j}) \neq d_{\text{ref}}(\mathbf{p}_w^{i,j})$ and $d(\mathbf{p}_w^{i,j}) = occ$).

The use of the criteria (1) to (3) has been introduced by the authors of [9]. The criterion (1) is calculated for each evaluation area. We present visual results with an error map:

- If the pixel is white, the correspondence is erroneous or bad.

- If the pixels is black it is a true negative.
- If the pixel is grey, the correspondence is correct.

The main advantage of this set of criteria, compared to the Scharstein and Szeliski protocol, is that it is more complete and it allows to evaluate precisely the efficiency in occlusion and discontinuity areas.

Whole Rank (WR) and Family Rank (FR) – Like Scharstein and Szeliski[4], methods are classified according to the mean of the ranks attributed to the measure for each evaluation criterion. Compared to their protocol, the number of criteria is more important in ours. In the tables, we note WR for indicating the whole rank of each correlation – compared to all the other correlation measures – and FR the rank into its own family – compared to the measures of the same family. In Table 14, these ranks are given for the *cones* pair, whereas in Table 15, they are presented for all the pairs: each rank is the mean of the ranks on all the images and each WR and FR are estimated on the mean of these mean ranks.

10 Experimental results

The measures presented in sections 3 to 8 have been tested and we used the bidirectional constraint that consists in estimating correspondences from left to right and then from right to left and in considering non-coherent matches as occluded pixels (these occluded pixels are shown in black in disparity maps).

In this section, we give some examples of the results but interested readers can find the details of the results for each of 42 images on this web site:

[4] http://cat.middlebury.edu/stereo/

http://perso.lcpc.fr/sylvie.chambon/correlationResults.html.

Influence of the size of the correlation window – We have tested window size between 3×3 and 15×15 to study the behavior of the measures with different window sizes. In Table 10, we give for each family, the mean of best window sizes, i.e. the size that permits the best values of the criterion, for each criterion on all the 42 tested images. A more complete version of this table can be found in the web page cited at the beginning of this section. We can notice that for the percentage of correct matches, the most efficient measures with a reasonable window size (7×7 or 9×9) are: the measures of family CROSS, ZD_1, LD_1, VD, GC, CENSUS, MAD and $R_{1,2,4,5}$.

FAMILY	COR	ACC	BAD	ERR	FPO	FNE	WOA	OA	OIA	DA
CROSS	9	11	11	9	5	13	3	5	13	7
CLASSICAL	9	11	11	9	5	13	3	5	13	7
DERIVATIVE	15	11	7	13	11	15	5	9	15	15
NON-PARAMETRIC	13	11	7	13	9	15	5	7	15	11
ROBUST	13	11	9	13	9	15	5	7	15	11

Table 10
Mean of best window sizes for each criterion ((W)OA: (Whole) Occlusion Area, OIA: Occlusion Influence Area, DA: Discontinuity Area) on all the 42 tested images and for each family.

The smaller the window, the more difficult to distinguish two different neighbourhoods and, moreover, the larger the window, the more possible to take into account pixels with different disparities. Consequently, the number of erroneous matches in the occlusion areas increases as the window size increases and, as expected, best results in Whole Occlusion Area (WOA), Occlusion Area (OA) and DA (Discontinuity Area) are obtained with small windows, lower than 9×9, for most of the measures. Moreover, Figure 7 illustrates the fact that for a 9×9 window, the percentage of correct matches is optimal for five selected measures (one of the best

measures for each family) and that the smaller the window, the better the results in the WOA. These two results are in contradiction and with a fixed size strategy, it seems difficult to select the most suitable size of windows.

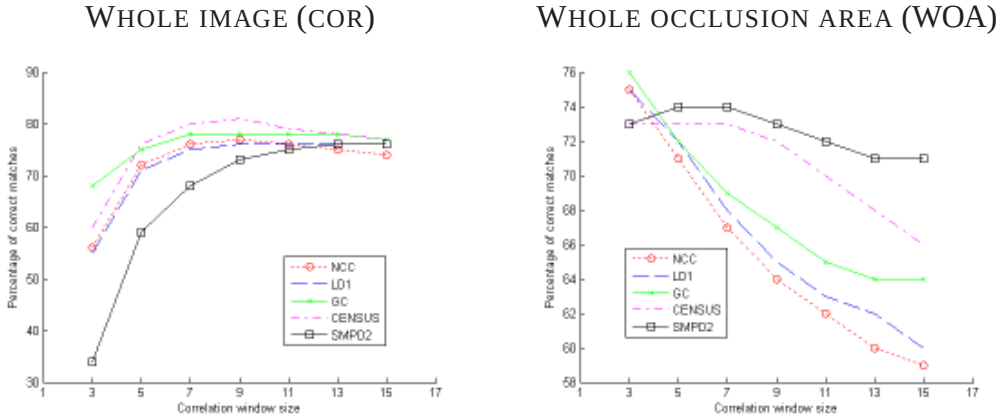


Fig. 7. Variation of the mean of two criteria over all the images for five measures – We show the results only for the most significant measures for each family, i.e. measures with rank 1 for the global results, cf. Table 15. The left graph represents the percentage of correct matches in the whole image whereas the right graph illustrates the percentage of good results in the WOA. The robustness of $SMPD_2$ is illustrated by its results in the WOA.

However, in order to reduce the size of data to analyse, we decided to focus on the results obtained with the percentage of correct matches in the whole image, COR, and in the WOA (Whole Occlusions Area), cf. Figure 7. It seems reasonable to compare the results with a 9×9 window for each similarity measure that is evaluated, because, the percentage COR is the best (or near the best) and the results in WOA are quite good (in most of the cases). Clearly, it corresponds to maximizing the correct matches in the whole image, for classical measures, without degrading (too much) the efficiency in WOA. Moreover, with a bigger window, the percentages of correct matches are not improved a lot for the other kind of measures, whereas, the percentages of correct matches in WOA are significantly decreased. This size coincidentally corresponds to maximizing erroneous matches, cf. Table 10. Moreover, this size is also a standard for this data set in the literature [40,44].

Execution time – The first three families have reasonable execution times whereas

the two last ones are expensive, cf. Table 11. Our implementation is not optimal, however, κ, χ, CENSUS and RZNCC, even optimised, are very expensive.

MEASURE	TIME	MEASURE	TIME	MEASURE	TIME	MEASURE	TIME	MEASURE	TIME
NIS	2.38	PRATT	5.33	ME_1	6.25	ME_2	33.06	R_4	43.64
VAD	2.77	ZD_1	5.23	OCM	6.37	MAD	33.23	ME_5	49.48
NCC	3.09	SEK_1	4.83	ISC	6.67	R_1	35.5	$\mathrm{D}_{0.1}$	74.88
MOR	3.91	ZNCC	5.4	ME_7	11.0	ZNCC_R	35.55	JEFF	83.84
SES_1	4	LD_1	5.44	SCC	13.02	R_2	38.1	QUAD	91.35
D_1	4.13	ME_4	5.89	LMP_2	16.65	R_5	38.33	κ	167.27
NA_1	4.57	χ^2	5.93	GC	17.6	ME_8	39.46	χ	179.82
NA_2	4.61	ME_6	5.94	ME_3	24.1	SMPD_2	39.77	CENSUS	305.67
RANK_1	4.73	K_4	6.1	LTP_2	25.44	R_3	40.44	RZNCC	725.7

Table 11
Execution times (in seconds, for the image of Figure 6, with a 9×9 window with a processor intel®core™2 duo of 2 GHz) – Measures are classified according to growing execution time. It gives only some indications because the implementation (for some of them, like SAD) has not been optimized and, for example, box-filter techniques [45,46] can be used.

Gaussian and impulsive noises – For Gaussian noise, the lower the signal-to-noise ratio, the larger the window to obtain 100% correct matches. The CROSS and CLASSICAL families are robust against Gaussian noise. In the DERIVATIVE family, only GC is efficient, and some measures of the NON-PARAMETRIC and ROBUST families are not efficient: RANK_P, the partial correlation, QUAD and MAD. The most robust measures against Gaussian noise are: SSD, ME_1 and ME_7. The study of results obtained in the presence of impulsive noise gives a first approximation of the behaviour of the measures with occlusions: damaged pixels can be considered as occluded pixels. The CROSS and CLASSICAL families are not robust. The measures of the DERIVATIVE family have a good behaviour except PRATT. Two problems occur: first, the PRATT convolution, with these two images, gives an image with repetitive patterns and, second, it is highly sensitive to only one erroneous pixel. The NON-PARAMETRIC family is effective but, the larger the proportion of noise,

the larger the correlation window to obtain good results. The ROBUST family is, as expected, the most efficient. Some results are in Table 12 where we give the smallest size of windows for having good performances with Gaussian and impulsive noises with one of the best measures for each family (for the robust one, two measures: a known measure and a proposed one).

	GAUSSIAN, SNR = 0.1		GAUSSIAN, SNR = 0.2		IMPULSIVE, 5%		IMPULSIVE, 10%	
ZNCC	3	5	21	9	7	25*	9	25*
D_1	7	5	21	9	7	23	5	25*
GC	7	5	25	9	7	3	9	25*
ISC	11	7	25	13	9	5	9	9
$D_{0.1}$	11	7	25	25	3	5	3	5
ME_5	5	5	25	15	3	3	3	5

Table 12
Results for *random* (first columns) and *sand* (second columns) presented in Figure 4 (SNR: Signal to Noise Ratio) – We present the smallest size of the window necessary to obtain 100% of correct matches. Tested window sizes vary between 3×3 and 25×25. In some cases, denoted by *, the measure never reaches this performance.

Performance analysis – In Table 13, we present the averages and the variances of the results on the 42 pairs with the best measure for each family. Variances are high for correct matches, it is around 14%. It illustrates how the data set is heterogeneous. Moreover, the results for the *Cones* pair are detailed in Table 14 and Figure 9. These images are less difficult than the others because there are no low textured areas and the occlusion areas are numerous but small. For these reasons, the CROSS and CLASSICAL families perform better than with the other images.

In Figure 8, we present the variations of the percentage of correct matches on the whole image and on the Whole Occlusion Area (WOA), for the 42 images, for five significant measures, with a 9×9 window. It illustrates the matching difficulties in all the images (the set of 2006 contains images with important illumination changes, low textured areas, complex scenes), and, it demonstrates that $SMPD_2$ is very efficient in the WOA and close to the best measures in the whole image.

MEASURE	COR	ACC	BAD	ERR	FPO	FNE	WOA	OA	OIA	DA
NCC	78 (12)	2 (2)	0 (0)	2 (3)	1 (1)	13 (8)	65 (11)	80 (17)	45 (10)	54 (11)
LD_1	78 (13)	2 (2)	0 (1)	2 (4)	1 (1)	13 (8)	66 (11)	80 (17)	48 (10)	57 (12)
GC	80 (12)	2 (3)	0 (0)	3 (5)	1 (1)	**11** (7)	68 (11)	78 (17)	53 (10)	61 (12)
CENSUS	**81** (12)	2 (3)	0 (0)	2 (2)	1 (1)	12 (8)	71 (9)	84 (10)	55 (13)	**67** (10)
$SMPD_2$	74 (15)	**1** (2)	0 (0)	**1** (3)	**0** (1)	19 (13)	**73** (11)	**85** (14)	**57** (13)	64 (14)

Table 13
Averages and variances (in parentheses) of the percentages obtained for most of the criteria with different scenes ((W)OA: (Whole) Occlusion Area, OIA: Occlusion Influence Area, DA: Discontinuity Area) with 9×9 correlation window for the 42 pairs. Bold letters correspond to the best results over the 5 presented results.

MEASURE	COR	ACC	BAD	ERR	FPO	FNE	WOA	OA	OIA	DA	WR
ZNCC	81.08 [12]	1.12 [21]	0.55 [20]	2.67 [21]	3.83 [39]	10.75 [6]	63.59 [23]	72.63 [39]	57.30 [17]	59.46 [12]	18
LD_1	81.84 [8]	0.98 [23]	0.47 [26]	2.14 [12]	3.42 [26]	11.14 [10]	64.58 [19]	75.56 [27]	56.94 [19]	58.97 [16]	12
GC	82.66 [4]	1.23 [17]	0.5 [24]	2.54 [19]	4.09 [44]	8.98 [5]	72.87 [4]	70.74 [42]	74.36 [3]	80.24 [2]	3
ISC	82.87 [3]	0.37 [40]	0.31 [36]	2.34 [16]	3.18 [21]	10.92 [10]	73.79 [2]	77.26 [19]	71.37 [4]	76.81 [4]	2
$SMPD_2$	**85.86** [1]	0.46 [39]	0.2 [42]	**1.22** [5]	**2.91** [19]	9.07 [6]	**77.4** [1]	**79.2** [17]	**76.14** [1]	**78.87** [3]	1

Table 14
Results achieved with the *cones* pair ((W)OA: (Whole) Occlusion Area, OIA: Occlusion Influence Area, DA: Discontinuity Area, WR: Whole Rank) – We give the results of the best correlation measure in each family (in brackets, the WR on the criterion). The complete results, i.e. are available in Table 18. We specify the rank for the results of each measure over all the measures (WR). If we compare with the results on the 42 images, cf. Table 15, the best measures of the first and the fourth families are not the same (NCC, CENSUS).

Table 15 summarizes the results for the best measures. For each criterion, we present the rank based on the mean of the ranks on all the tested images, then the classification per family (FR) and the global rank (for all the criteria, WR). More details can be found on our web page [5] .

We present a global analysis of the results:

[5] http://perso.lcpc.fr/sylvie.chambon/correlationResults.html

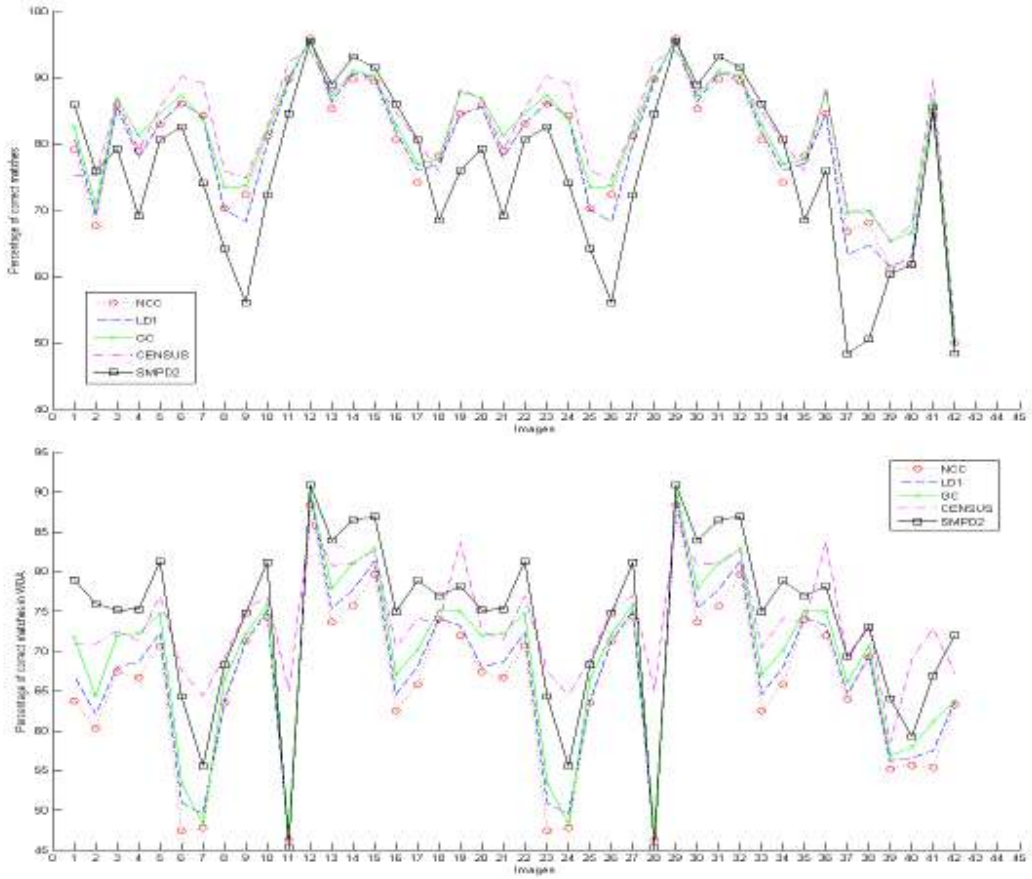


Fig. 8. Variations of percentage of correct matches on all the images for the best measure per family – The numbers of the images correspond to the order of the presentation in the § **tested images**, section 9. The first graph illustrates the variation on the whole image whereas the second one only concerns the correct matches on the Whole Occlusion Area (WOA). These graphs illustrate how the results can be different from an image to another and also that $SMPD_2$ is most of the time the best (30/42 images) in the WOA whereas it is not the case for the percentage of correct matches in the whole image.

- For each family:
  - · CROSS – NCC is the best but in fact, these measures hold similar results, and belong to the ten best measures for the percentage of correct matches.
  - · CLASSICAL – LD_1 (LSAD) is the best even if ZD_1 (ZSAD) reaches the best percentage of correct matches. These measures have an interesting rank except K_4 which was designed to be insensitive to Gaussian noise but is highly sensitive to outliers (powers of 4 and 2 are used).
  - · DERIVATIVE – GC is the best and gives the best percentage of correct matches and false negatives. The other measures are not well ranked.

- NON-PARAMETRIC – CENSUS is the best. RANK_1 is the second one and reaches a good percentage of correct matches in the occlusion influence area and the discontinuity area. We can also notice the ISC measure that gives interesting results. In particular, it is the best measure of the family for the results obtained with *cones* pair.
- ROBUST – SMPD_2 is the best and it obtains the best percentage of correct matches in the whole occlusion area. LTP_2 also furnishes good results.

- In conclusion of this evaluation, the best measure is CENSUS, and except it, GC and RANK_1, the ten best measures belong to the ROBUST family. As expected, SMPD_2, which is designed for, obtains the best results in the whole occlusion area.
- In general, best measures in non-occluded areas, like NCC, do not perform well in the whole occlusion area, and, the best measures in the whole occlusion area, like SMPD_2, do not achieve good results with non-occluded areas.

For reason of space, we cannot show all the results for each pair, however, we can conclude that even if CENSUS is the first measure in our evaluation, it is not always the best measure on each image separately. Moreover, SMPD_2 is not always the first but it always belongs to the ten best measures, except with one pair, *plastic* (rank 18) which possesses large low textured areas, and SMPD_2 is not efficient with these areas. In the same way, CENSUS is not always in the ten best measures, and in six images highly textured, *cloth 1* (rank 28), *cloth 3* (rank 18), *cloth 4* (rank 19), *flowerpots* (rank 24), *stereo* (rank 42) and *wood 2* (rank 13). In fact, in this kind of images, CENSUS gives a lot of false positives.

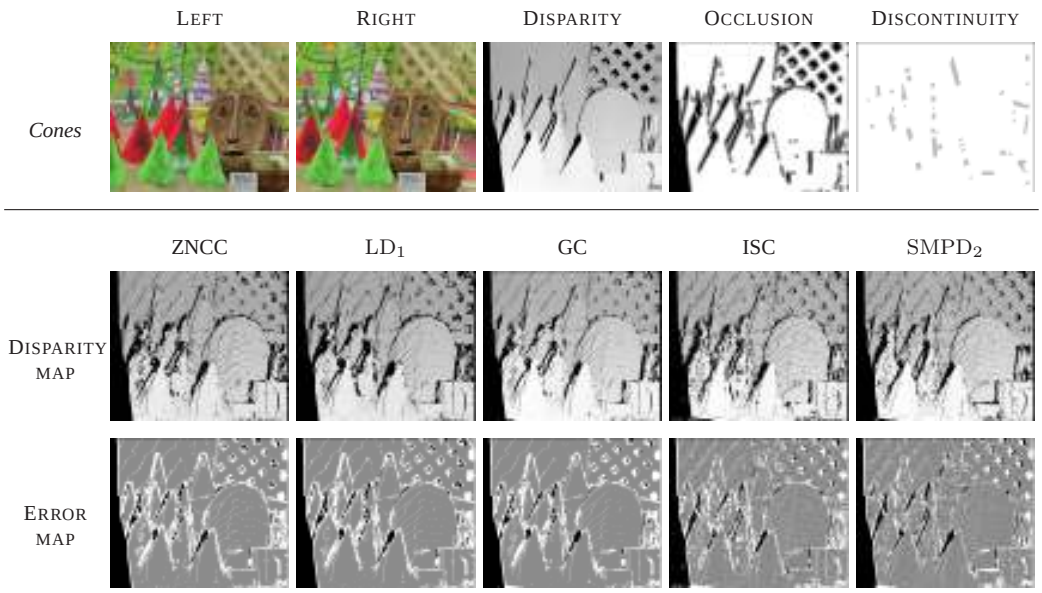


Fig. 9. Maps obtained with the *cones* pair – Differences between the disparity maps are not very significant. However, we can notice the good results of SMPD_2 in the occlusion areas.

11 Discussion

The results illustrate the interesting performances of CENSUS and SMPD_2 but we have also noticed that the ten best measures include other of the proposed robust measures: LTP_2, ME_3, R_1, R_3, R_4 and R_5. Compared to SMPD_2, the main advantage of the R-estimator measures is that they outperform SMPD_2 in the non-occluded areas (with better results than classical measures in occluded areas and less false positives than the ten best other measures). In consequence, if it is important to obtain acceptable performances in these two different areas, R-estimator measures are well suited and R_1 has the smallest execution time. If the execution time is not critical, CENSUS obtains better results than R_1 both in occluded and non-occluded areas. In fact, it seems that CENSUS is more interesting than SMPD_2 with low textured images whereas SMPD_2 is more interesting than CENSUS in well textured images. It can be explained by this fact: CENSUS allows to use more information (80 bits for 9×9 window) than SMPD_2 (only 8

bits for a gray level difference) for each pixel. In consequence, in low textured areas, the CENSUS descriptor is more discriminant than the gray level used in SMPD_2 to find the right correspondences. It is the reverse in well textured areas. This remark is corroborated by the individual results on each image. We have identified the images for which each measure gives the best results, i.e. obtains WR$= 1$ (see Scharstein and Szeliski website for more details about the images `http://cat.middlebury.edu/stereo/`):

- **CENSUS (14 images)**: barn2, books, bowling1, bull, *journaux*, *livres*, laundry, lampshade1, lampshade2, midd2, poster, plastic, reindeer, venus ;
- **SMPD_2 (13 images):** art, baby2, cloth1 to cloth4, cones, dolls, moebius, monopoly, rocks2, sawtooth, *stereo*;
- **RANK_1 (6 images):** aloe, baby1, baby3, bowling2, rocks1, wood1;
- **LTP_2 (5 images):** barn1, map, *murs*, teddy, tsukuba;
- **GC:** wood2;
- **ME_3:** midd1, *plante*;
- **LD_1:** flowerpots.

The 13 images where SMPD_2 has obtained the best results represent highly complex and textured scenes with a lot of occlusions, which confirms its superiority in this case. When no intensity change is observed between the two images, it is more appropriate to use the non-centered version: LTP_2. The CENSUS and the RANK_1 measures are the most efficient in the case of scenes with low textured areas and radiometric distortions (it is one of the properties of this measure which has been introduced to be independent from the magnitude of the intensities). In very difficult images, with high ambiguities (due to low textured areas), like flowerpots, it is hard to conclude which measure is well adapted but it seems that M-estimator would be an interesting choice.

Family	Measure	Cor	Acc	Bad	Err	FPo	FNe	WOA	OA	OIA	DA	WR	FR
Cross	NCC	5	25	23	10	31	8	31	32	27	22	23	**1**
	ZNCC	13	31	31	13	36	13	27	38	19	19	30	3
Classical	D_1	23	11	12	29	25	20	19	27	15	17	19	3
	LD_1	6	26	22	16	22	7	24	26	20	13	13	**1**
Derivative	**GC**	4	19	20	20	38	3	10	39	9	3	8	**1**
Non-parametric	$RANK_1$	8	19	34	12	35	9	5	36	3	3	7	2
	CENSUS	**1**	17	37	5	38	2	2	11	2	**1**	**1**	**1**
Robust	$\mathbf{RM_{LMS}}$	2	38	41	2	2	29	3	2	7	12	6	5
	$\mathbf{LTP_2}$	29	6	7	34	11	26	4	12	5	10	3	2
	$\mathbf{SMPD_2}$	15	34	38	3	7	17	**1**	6	**1**	2	2	**1**
	$\mathbf{ME_3}$	26	9	9	32	14	22	6	17	6	6	5	4
	$\mathbf{R_3}$	4	29	27	11	20	5	14	23	8	5	4	3
	$\mathbf{R_4}$	10	33	32	9	19	12	18	22	11	8	10	7
	$\mathbf{R_5}$	12	28	25	20	21	11	16	25	9	7	9	6

Table 15
Classification ((W)OA: (Whole) Occlusion Area, OIA: Occlusion Influence Area, DA: Discontinuity Area) – We summarize the results over the 42 image pairs for the ten best measures and also the best measure for each family. Moreover, we add the results of well known measures to compare our work to the state of the art: ZNCC and D_1 (SAD). The complete results can be find here: http://perso.lcpc.fr/sylvie.chambon/correlationResults.html.
For each evaluation criterion, the rank is estimated with the mean of the ranks obtained with all the images. The Whole Rank (WR) is estimated by comparing the mean of these mean ranks on all the correlation measures whereas the Family Rank (FR) takes into account only the measures of the same family.

Finally, a summary is given in tables 16 and 17 in order to propose keys to choose the most adapted measure for each application. In Table 16, we show performances against execution time: it appears that the measure RANK_1 offers the best compromise. This measure can be used if the presence of false positives is not important for the application (high percentage of false positive is the drawback of this measure). In Table 17, we summarize performances in occluded areas against non-occluded areas: it reveals that NCC, GC, RANK_1, R_1 to R_4, and CENSUS are good compromises to obtain both acceptable results in occluded and non-occluded areas.

T \ R	BAD	ACCEPTABLE	GOOD
HIGH	JEFF	QUAD, κ, χ	RM_{LMS}, CENSUS
ACCEPTABLE	$ZNCC_R$	SCC, $D_{0.1}$, LMP_2, ME_2, ME_5, ME_7, ME_8, MAD, R_1, R_2	GC, LTP_2, R_5, $SMPD_2$, ME_3, R_3, R_4
SHORT	NIS, SES_1, SEK_1, PRATT, K_4, OCM, ME_6	ZNCC, NCC, MOR, D_1, ZD_1, LD_1, VAD, χ_2, ISC, ME_1, ME_4	$RANK_1$

Table 16
Summary of the performances (**R**: rank) versus the execution time (**T**) – Time is short when it is less than 10 s, it is acceptable when less than 1 minute and high over. Rank is good if the measure is in the top 10 and bad in the last 10.

Occ. \ Non Occ.	BAD	ACCEPTABLE	GOOD
BAD	K_4, SES_1, SEK_1, NIS, OCM, PRATT, JEFF, $ZNCC_R$, ME_6	ZNCC, VAD, χ_2	MOR
ACCEPTABLE	LMP_2, ME_5	D_1, ZD_1, LD_1, ISC, SCC, κ, χ, $D_{0.1}$, LTP_2, ME_1, ME_2, ME_3, ME_4, ME_7, ME_8, R_5	**NCC**, **GC**, $RANK_{\mathbf{1}}$, $R_{\mathbf{1}}$, $R_{\mathbf{2}}$, $R_{\mathbf{3}}$, $R_{\mathbf{4}}$, **CENSUS**
GOOD	MAD	$RM_{\mathbf{LMS}}$, **QUAD**, $SMPD_{\mathbf{2}}$	

Table 17
Summary of the performances in non-occluded areas (**Non Occ**) versus the performances in occluded areas (**Occ**) – Result in non-occluded areas is good if the ranks for correct matches and false negatives, cf. Table 15, are better than 10 and bad if they are in the last 10. Result in occluded areas is good if the ranks for correct matches in the whole occlusion areas and false positives are better than 10 and bad if they are in the last 10. In bold, it correspond to the best compromises to obtain good results both in occluded areas and non-occluded areas.

12 Conclusion

In the context of similarity matching, we have presented a review of the correlation measures: forty measures have been detailed and classified into five families including six types of correlation measures based on robust statistics tools previously proposed in order to take into account the occlusion problem. We have set up an evaluation and comparison protocol adapted to study the behaviour of correlation-based methods and, in particular near occlusion and discontinuity areas. The results highlight the best measure near occlusion: $SMPD_2$ and near discontinuity: CENSUS. It also demonstrates the drawbacks of the robust measures: the difficulties

with non-occluded areas. This behavior is coherent with the definition: it allows the method to be insensitive to large differences induced by an occlusion and, in consequence, it is ambiguous in non-occluded regions, because, many candidates can obtain the maximal correlation score (the true correspondent but also many wrong correspondents). The discussion about all the results tries to give some advices about which robust measure should be used for a given application. First, R-estimator is the most interesting when acceptable results are needed in occluded areas but also in non-occluded areas (these measures are not as efficient as SMPD_2 in occluded areas but they are both better in occluded and non-occluded areas than classic measures). Second, M-estimator measures seem to be promising with very ambiguous scenes (low textured with occlusions). Finally, CENSUS is the first one but detailed results highlight the fact that it is not the most efficient in highly textured images and it can present a lot of false positives.

An extension of this work will include the study of a more efficient measure that should combine the advantages of a robust measure, like SMPD_2, and a classical measure, GC or CENSUS. To do this work, a study of the complementarity of these measures is needed. In particular, we will evaluate the different areas in the images where each measure can give, alone, the true correspondent. Then, it seems difficult to propose a single measure that can have both the advantages of the two measures (because the definitions are incompatible) but we hope to improve the performances by using the two measures together (and maybe more measures to be more robust). An algorithm based on merging disparity maps has been implemented and gives encouraging results. Our work will focus on this kind of approach.

References

[1] S. Chambon, A. Crouzil, Dense matching using correlation: new measures that are robust near occlusions, in: British Machine Vision Conference, BMVC, Vol. 1, Norwich, United Kingdom, 2003, pp. 143–152.

[2] L. G. Brown, A Survey of Image Registration Techniques, ACM Computing Surveys 24 (4) (1992) 325–376.

[3] J. Puzicha, T. Hofmann, J. Buhmann, Non-parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, San Juan, Porto Rico, 1997, pp. 267–272.

[4] J. Mulligan, V. Isler, K. Daniilidis, Performance of Stereo for Tele-presence, in: IEEE International Conference on Computer Vision, ICCV, Vol. 2, Vancouver, Canada, 2001, pp. 558–565.

[5] D. Garcia, J. Orteu, L. Penazzi, A combined Temporal Tracking and Stereo-correlation Technique for Accurate Measurement of 3D Displacements: Application to Sheet Metal Forming, Journal of Materials Processing Technology 125–126 (1) (2002) 736–742.

[6] S. Porter, M. Mirmehdi, B. Thomas, Video Indexing using Motion Estimation, in: British Machine Vision Conference, BMVC, Vol. 2, Norwich, United Kingdom, 2003, pp. 659–668.

[7] P. Aschwanden, W. Guggenbül, Experimental results from a comparative study on correlation type registration algorithms, in: W. Förstner, S. Ruwiedel (Eds.), Robust computer vision: Quality of Vision Algorithms, Wichmann, Karlsruhe, Germany, 1992, pp. 268–282.

[8] A. Giachetti, Matching techniques to compute image motion, International Journal of

Image and Vision Computing 18 (3) (2000) 245–258.

[9] Z. Lan, R. Mohr, Robust location based partial correlation, Research report RR-3186, Institut National de Recherche en Informatique et en Automatique, INRIA (June 1997).

[10] Y. Rubner, J. Puzicha, C. Tomasi, J. Buhmann, Empirical Evaluation of Dissimilarity Measures for Color and Texture, Computer Vision and Image Understanding, CVIU 84 (1) (2001) 25–43.

[11] K. Yoon, I. Kweon, Locally Adaptive Support-Weight Approach for Visual Correspondence Search, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, Vol. 2, San Diego, United States, 2005, pp. 924–931.

[12] C. Jawahar, P. Narayanan, Generalised correlation for multi-feature correspondence, The Journal of the Pattern Recognition Society, PR 35 (6) (2002) 1303–1313.

[13] H. Hirschmüller, D. Scharstein, Evaluation of stereo matching costs on images with radiometric differences, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 31 (9) (2009) 1582–1599.

[14] T. Kanade, M. Okutomi, A Stereo Matching Algorithm with an Adaptive Window: Theory and experiment, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 16 (9) (1994) 920–932.

[15] A. Fusiello, V. Roberto, E. Trucco, Efficient stereo with multiple windowing, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, San Juan, Porto Rico, 1997, pp. 858–863.

[16] V. Ferrari, T. Tuytelaars, L. Van Gool, Wide-baseline Multiple-view Correspondences, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, Vol. 2, Madison, United States, 2003, pp. 718–725.

[17] H. Moravec, Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover, Thesis, University of Carnegie Mellon, CMU, Pittsburgh, United States

(September 1980).

[18] G. Cox, Template Matching and Measures of Match in Image Processing, Research report, University of Cape Town, South Africa (July 1995).

[19] M. Rziza, D. Aboutajdine, Dense disparity map estimation using cumulants, in: Conference on Telecommunications, ConfTele, Figueira da Foz, Portugal, 2001.

[20] T. Kanade, A. Yoshida, K. Oda, H. Kano, M. Tanaka, A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, San Francisco, United States, 1996, pp. 196–202.

[21] M. Okutomi, T. Kanade, A Multiple-Baseline Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 15 (4) (1993) 358–363.

[22] P. Seitz, Using local orientational information as image primitive for robust object recognition, in: Visual Communication and Image Processing IV, Vol. SPIE–1199, 1989, pp. 1630–1639.

[23] H. Nishihara, PRISM, a pratical real-time imaging stereo matcher, Technical report A. I. Memo 780, Massachusetts Institute of Technology, MIT, United States (1984).

[24] M. Nack, Temporal registration of multispectral digital satellite images using their edge images, in: Astrodynamics Specialist Conference, Nassau, Bahamas, 1975, pp. AAS75–104.

[25] W. K. Pratt, Digital image processing, Wiley-Interscience Publication, New-York, United States, 1978, Ch. 20, pp. 666–667.

[26] F. Ullah, S. Kaneko, S. Igarashi, Orientation Code Matching For Robust Object Search, Proceedings of IEICE Transactions on Information and Systems E-84-D (8) (2001) 999–1006.

[27] A. Crouzil, L. Massip-Pailhes, S. Castan, A New Correlation Criterion Based on Gradient Fields Similarity, in: IEEE International Conference on Pattern Recognition, ICPR, Vol. 1, Vienna, Austria, 1996, pp. 632–636.

[28] S. Kaneko, I. Murase, S. Igarashi, Robust Image registration by Increment Sign Correlation, The Journal of the Pattern Recognition Society, PR 35 (10) (2002) 2223–2234.

[29] S. Kaneko, Y. Satoh, S. Igarashi, Using selective correlation coefficient for robust image registration, The Journal of the Pattern Recognition Society, PR 36 (5) (2003) 1165–1173.

[30] R. Zabih, J. Woodfill, Non-parametric Local Transforms for Computing Visual Correspondence, in: European Conference on Computer Vision, ECCV, Stockholm, Sweden, 1994, pp. 151–158.

[31] D. Bhat, S. Nayar, Ordinal Measures for Image Correspondence, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 20 (4) (1998) 415–423.

[32] P. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, J. Wiley & Sons, New-York, United States, 1987.

[33] P. Huber, Robust statistics, J. Wiley & Sons, New-York, United States, 1981, Ch. 8, pp. 204–205.

[34] M. Trujillo, E. Izquierdo, A robust correlation measure for correspondence estimation, in: International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT, Thessaloniki, Greece, 2004, pp. 155–162.

[35] J. Delon, B. Rougé, Analytic study of the stereoscopic correlation, Research report 2004-19, Centre de Mathématique et de Leurs Applications, CMLA, (Ecole Normale Supérieure de Cachan, ENS) (2004).

[36] P. Rousseeuw, C. Croux, L_1-Statistical Analysis, in: Explicit Scale Estimators with High Breakdown Point, Elsevier, Amsterdam, Holland, 1992, pp. 77–92.

[37] Z. Zhang, Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting, Research report RR-2676, Institut National de Recherche en Informatique et en Automatique, INRIA (October 1995).

[38] Y. Wang, D. Wiens, Optimal, robust R-estimators and test statistics in the linear model, Statistics and Probability Letters 14 (3) (1992) 179–188.

[39] D. Wiens, J. Zhou, Bounded-influence rank estimation in the linear model, The Canadian Journal of Statistics 22 (2) (1994) 233–245.

[40] D. Scharstein, R. Szeliski, A taxomomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision, IJCV 47 (1) (2002) 7–42.

[41] D. Scharstein, R. Szeliski, High-Accuracy Stereo Depth Maps Using Structured Light, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, Vol. 1, Madison, United States, 2003, pp. 195–202.

[42] D. Scharstein, C. Pal, Learning conditional random fields for stereo, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, Minneapolis, United States, 2007.

[43] H. Hirschmüller, D. Scharstein, Evaluation of cost functions for stereo matching, in: IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, Minneapolis, United States, 2007.

[44] H. Hirschmüller, P. Innocent, J. Garibaldi, Real-time correlation-based vision with reduced border errors, International Journal of Computer Vision, IJCV 47 (1–3) (2002) 229–246.

[45] O. Faugeras, B. Hotz, Z. Zhang, P. Fua, Real time correlation-based stereo: Algorithm, implementation and applications, Research report RR-2013, Institut National de Recherche en Informatique et en Automatique, INRIA (August 1993).

[46] H. Mayer, Analysis of Means to Improve Cooperative Disparity Estimation, in: Proceedings of the ISPRS Workshop on Photogrammetric Image Analysis, Technical University Munich, Germany, 2003, pp. 25–31.

FAMILY	MEASURE	COR	ACC	BAD	ERR	FPO	FNE	WOA	OA	OIA	DA	WR	FR
CROSS	NCC	80 (16)	1 (19)	0 (21)	2 (15)	3 (29)	11 (13)	62 (29)	75 (30)	53 (28)	55 (23)	24	2
	ZNCC	81 (13)	1 (21)	0 (20)	2 (21)	3 (38)	10 (7)	63 (24)	72 (38)	57 (18)	59 (13)	19	1
	MOR	80 (15)	1 (18)	0 (18)	2 (23)	3 (41)	10 (5)	63 (26)	71 (41)	56 (22)	59 (16)	26	3
CLASSIC	D_1	74 (25)	1 (12)	0 (12)	3 (28)	3 (34)	16 (24)	64 (19)	74 (33)	57 (17)	55 (24)	27	4
	ZD_1	81 (8)	1 (22)	0 (25)	2 (13)	3 (31)	11 (9)	64 (22)	75 (31)	56 (19)	59 (15)	15	2
	LD_1	81 (9)	0 (23)	0 (26)	2 (11)	3 (26)	11 (11)	64 (20)	75 (26)	56 (20)	58 (17)	13	1
	VAD	79 (19)	1 (16)	0 (17)	2 (16)	3 (23)	12 (17)	60 (32)	75 (23)	49 (32)	52 (28)	23	3
	K_4	50 (38)	2 (7)	1 (7)	10 (41)	3 (21)	31 (37)	48 (42)	77 (21)	27 (41)	30 (40)	40	5
DERIVATIVE	SES_1	55 (37)	0 (25)	0 (35)	2 (19)	1 (2)	39 (42)	53 (39)	90 (2)	27 (42)	30 (41)	38	2
	SEK_1	56 (36)	0 (27)	0 (34)	5 (35)	2 (3)	34 (39)	52 (40)	83 (3)	30 (39)	34 (39)	41	4
	NIS	18 (43)	0 (43)	0 (39)	5 (34)	0 (1)	74 (43)	40 (43)	93 (1)	3 (43)	4 (43)	43	6
	OCM	71 (28)	0 (32)	0 (31)	4 (32)	3 (22)	19 (28)	58 (35)	76 (22)	45 (35)	51 (30)	39	3
	PRATT	65 (32)	0 (42)	1 (9)	7 (38)	3 (19)	22 (31)	54 (38)	77 (19)	38 (38)	44 (34)	42	5
	GC	82 (5)	1 (20)	0 (23)	2 (18)	4 (42)	8 (3)	67 (8)	70 (42)	64 (5)	65 (5)	8	1
NON-PARAMETRIC	JEFF	67 (31)	2 (8)	1 (8)	5 (33)	3 (37)	19 (27)	61 (31)	73 (37)	52 (29)	51 (31)	36	8
	χ_2	77 (21)	1 (15)	0 (16)	2 (24)	3 (36)	13 (18)	62 (27)	74 (36)	54 (27)	55 (26)	33	7
	ISC	83 (4)	0 (40)	0 (36)	2 (14)	3 (17)	10 (8)	69 (5)	78 (17)	63 (6)	65 (6)	4	1
	SCC	79 (20)	1 (17)	0 (19)	2 (22)	3 (39)	12 (16)	62 (28)	72 (39)	56 (23)	57 (18)	30	6
	$RANK_1$	83 (3)	0 (35)	0 (33)	3 (25)	4 (43)	7 (1)	66 (9)	65 (43)	67 (3)	68 (3)	16	5
	CENSUS	84 (2)	0 (34)	0 (32)	2 (12)	3 (40)	8 (2)	70 (3)	71 (40)	69 (2)	69 (2)	7	3
	κ	80 (17)	0 (36)	0 (37)	1 (4)	2 (14)	14 (21)	68 (7)	80 (14)	60 (8)	60 (9)	6	2
	χ	77 (22)	0 (30)	0 (38)	1 (6)	2 (5)	17 (26)	66 (10)	81 (5)	56 (24)	56 (20)	11	4
ROBUST	RM_{LMS}	81 (10)	0 (39)	0 (43)	0 (1)	2 (4)	14 (20)	71 (2)	82 (4)	64 (4)	66 (4)	2	2
	QUAD	81 (11)	0 (37)	0 (40)	1 (5)	2 (8)	13 (19)	68 (6)	81 (7)	59 (9)	60 (10)	3	3
	$ZNCC_R$	68 (30)	0 (33)	0 (30)	1 (7)	2 (12)	25 (34)	58 (36)	80 (12)	42 (36)	42 (37)	35	20
	$D_{0.1}$	57 (34)	4 (5)	2 (5)	7 (37)	3 (18)	24 (33)	62 (30)	78 (18)	51 (30)	48 (32)	32	18
	MAD	74 (26)	0 (41)	0 (42)	1 (2)	2 (10)	21 (30)	70 (4)	81 (10)	62 (7)	60 (7)	9	5
	LMP_2	56 (35)	1 (11)	0 (13)	3 (30)	2 (7)	34 (40)	63 (25)	81 (8)	50 (31)	45 (33)	28	15
	LTP_2	59 (33)	4 (6)	2 (6)	6 (36)	2 (15)	24 (32)	65 (13)	80 (15)	54 (26)	51 (29)	18	10
	$SMPD_2$	86 (1)	0 (38)	0 (41)	1 (3)	2 (16)	9 (4)	74 (1)	79 (16)	71 (1)	71 (1)	1	1
	ME_1	75 (23)	1 (14)	0 (15)	3 (26)	3 (35)	15 (22)	64 (17)	74 (35)	57 (13)	56 (21)	20	11
	ME_2	73 (27)	1 (10)	0 (11)	3 (29)	3 (32)	16 (25)	64 (16)	75 (32)	57 (15)	55 (25)	22	13
	ME_3	69 (29)	2 (9)	1 (10)	4 (31)	3 (20)	19 (29)	65 (11)	77 (20)	57 (12)	54 (27)	17	9
	ME_4	50 (39)	6 (4)	2 (4)	8 (39)	2 (9)	29 (35)	60 (34)	81 (9)	45 (33)	43 (35)	31	17
	ME_5	42 (41)	7 (1)	3 (2)	11 (42)	2 (11)	32 (38)	56 (37)	81 (11)	39 (37)	36 (38)	34	19
	ME_6	32 (42)	6 (2)	3 (1)	14 (43)	2 (13)	39 (41)	50 (41)	80 (13)	29 (40)	27 (42)	37	21
	ME_7	74 (24)	1 (13)	0 (14)	3 (27)	3 (33)	16 (23)	64 (18)	74 (34)	57 (16)	56 (22)	25	14
	ME_8	49 (40)	6 (3)	2 (3)	8 (40)	2 (6)	29 (36)	60 (33)	81 (6)	45 (34)	43 (36)	29	16
	R_1	81 (7)	0 (26)	0 (24)	2 (10)	3 (30)	11 (10)	64 (21)	75 (29)	56 (21)	59 (14)	14	8
	R_2	80 (18)	0 (24)	0 (22)	2 (20)	3 (28)	12 (14)	63 (23)	75 (28)	55 (25)	57 (19)	21	12
	R_3	82 (6)	0 (28)	0 (28)	2 (9)	3 (27)	10 (6)	65 (12)	75 (25)	58 (10)	60 (8)	5	4
	R_4	81 (14)	0 (31)	0 (29)	2 (8)	3 (24)	12 (15)	65 (15)	75 (24)	57 (14)	59 (12)	12	7
	R_5	81 (12)	0 (29)	0 (27)	2 (17)	3 (25)	11 (12)	65 (14)	75 (27)	57 (11)	60 (11)	10	6

Table 18

Percentages for each criterion and each measure obtained with the pair of images *cones* (with 9×9 correlation window size, (W)OA: (Whole) Occlusion Area, OIA: Occlusion Influence Area, DA: Discontinuity Area) – The Whole Rank (WR) is estimated by comparing, between each measure, the mean of the ranks on all the criteria whereas the Family Rank (FR) takes into account only the measures of the same family. Moreover, results in brackets correspond to the rank of the evaluation criterion compared to the other measures, for each criterion.