

Complex Correlation Statistic for Dense Stereoscopic Matching

Jan Čech and Radim Šára

Center for Machine Perception, Czech Technical University, Prague, Czech Republic
{cechj,sara}@cmp.felk.cvut.cz

Abstract

A traditional solution of area-based stereo uses some kind of windowed pixel intensity correlation. This approach suffers from discretization artifacts which corrupt the correlation value. We introduce a new correlation statistic, which is completely invariant to image sampling, moreover it naturally provides a position of the correlation maximum between pixels. Hereby we can obtain sub-pixel disparity directly from sampling invariant and highly discriminable measurements without any postprocessing of the discrete disparity map.

The key idea behind is to represent the image point neighbourhood in a different way, as a response to a bank of Gabor filters. The images are convolved with the filter bank and the complex correlation statistic (CCS) is evaluated from the responses without iterations. The magnitude of CCS measures the image similarity and the phase gives the sub-pixel position.

Our experiments show that CCS has even better sampling invariance and discriminability properties than the popular Birchfield-Tomasi dissimilarity, and the sub-pixel accuracy is higher than the Maximum Likelihood sub-pixel disparity window fitting disparity correction method we compared with. Results are shown on both synthetic and real data.

1 Introduction

In stereo, we have to recognize corresponding points, i.e. the image points which are the projection of the same spatial point, according to how much the image point neighbourhoods are similar, computing some kind of image correlation statistic.

A stereo algorithm usually consists of two essential modules: the measurement evaluation and the matching. The first process computes some kind of similarity (correlation) statistics between all potential correspondences. The second process takes this measurements and establishes matches according to some principle or optimizing some criterion. Both stages are important and dramatically influence the matching results, which could be seen in stereo evaluation works [11], [5].

This paper is exclusively devoted to the similarity evaluation stage introducing a new correlation statistic which can be used by various matching algorithms. We introduce a Complex Correlation Statistic (CCS) which is invariant to image sampling and allows sub-pixel match localization.

Birchfield and Tomasi [1] noticed that it is important the correlation statistic be invariant to image discretization and proposed a sampling-invariant pixel dissimilarity. It is a simple extension of Sum of Absolute Differences (SAD) based on a linear interpolation. It works quite well but it tends to fail where there are very high frequencies in the images. The aggregation of this pixel dissimilarity over a window has become popular in area-based stereo, e.g. [17].

Szeliski and Scharstein [15] continued with this effort and recommended several other matching scores based on interpolated image signals, e.g. to interpolate the images to a higher resolution, compute the usual statistics, aggregate and subsample to the original resolution. The drawback of this approach is that the increase in resolution is finite, which limits the discretization invariance property.

There are several possibilities to achieve sub-pixel matching. Again, the simplest way is to work with interpolated high resolution images. We could have the sub-pixel precision up to the level of interpolation and also the statistic will be less sensitive to image sampling. But, computational expenses increase dramatically.

A possible solution might be to interpolate in the space of correlation statistic (disparity space), like fitting a parabola between three values in the correlation table to find where the extreme of the statistic is. These methods were studied by Shimizu and Okutomi [12, 13], where they formulate which interpolant is suitable for which statistics, and proposed a method to compensate a systematic error using that method.

In theory, a sub-pixel disparity map could be also obtained from any probabilistic labeling framework, e.g. [14, 16], where the final disparity is determined as a mean value from all integer disparity labels. However, it has not been analyzed properly whether such an estimate is accurate.

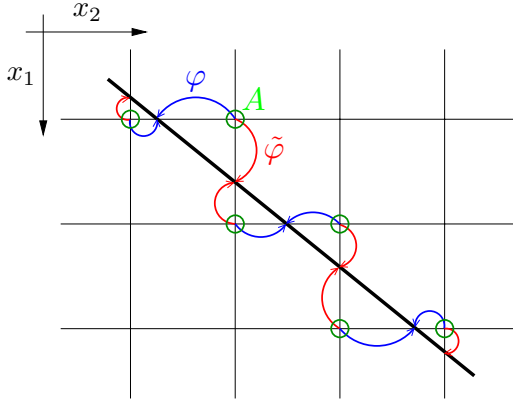


Fig. 1. Correlation table of the CCS



Fig. 2. Computational block

Sub-pixel matching precision can also be achieved using so called phased-based techniques. By the phase it is meant the Fourier phase of a signal or the angle of a response to a complex-valued filter in general. The method is based on the Fourier shifting theorem: a shift in the spatial domain corresponds to a shift of the phase in the frequency domain. Weng [18] introduced the windowed Fourier phase as a matching primitive in a hierarchical disparity estimation procedure. Sanger [7] used Gabor filters [2] and the final disparity is calculated as a weighted average of the disparities obtained from various filter bands. Fleet et al. [4, 3] also used Gabor filters. They introduced a stability criterion of the phase, which is based on the discrepancy between the tuning frequency of the Gabor filter and the local frequency of the response. Unstable estimates are not used in the subsequent iterative algorithm. Xiong and Shaffer [19, 20] proposed to use hypergeometric filters beside the Gabors. They use high-order Taylor expansion of the response which allows affine invariant estimation, but it requires solving a set of equations by an iterative procedure for each match.

Our approach is very close to phase-based techniques in the sense we model the image using Gabor filters and exploit the shifting theorem. But we do not estimate the disparity map directly, we embed the estimation of the sub-pixel disparity in the correlation statistic which is evaluated in a closed form without any complicated optimizations.

The paper is organized as follows. We will introduce our Complex Correlation Statistics in Sec. 2. The experiments will be presented in Sec. 3. Finally, Sec. 4 concludes the paper.

2 Complex Correlation Statistic

2.1 Definition

Let us have rectified images $I_L(x, y)$ and $I_R(x, y)$, in which epipolar lines coincide with image rows y in both images. Image similarities for all potential pixel correspondences $(x_1, y) \in I_L, (x_2, y) \in I_R$, for current scanline y form so called correlation table $c(x_1, x_2)$.¹ We define the Complex Correlation Statistic to be a complex number

$$CCS(x_1, x_2) = Ae^{j\varphi}, \quad (1)$$

where the magnitude A is the similarity value which is invariant to image sampling, and the phase φ shows the correct position of the match between pixels, see Fig. 1. The thick black line represents the truth matching, i.e. an image of a surface in the disparity space. Green circles mark cells of the correlation table at locations, where magnitude A should be the highest (ideally 1). Blue arrows represent the angle φ pointing towards the correct match position in the horizontal direction. Red arrows are pointing towards the correct match position in the vertical direction. This is the angle $\tilde{\varphi}$ of the complementary correlation $\tilde{C}CS(x_1, x_2) = Ae^{j\tilde{\varphi}}$. Magnitudes of CCS and $\tilde{C}CS$ are the same, the only difference is in phases. These statistics are evaluated in each cell of the correlation table. A flowchart of this procedure is sketched in Fig. 2. The inputs are intensity values of the neighbourhood \mathcal{N} of the left I_L and right I_R image at position x_1, x_2 , respectively on the y th row. Swapping the inputs f, g of this block causes swapping the CCS to $\tilde{C}CS$.

In the following sections, we will describe what is inside the CCS-block. First, we give the formula and then we explain it.

¹For the case of a standard windowed statistic, e.g. SAD, it contains the sum of absolute differences of pixel intensities in windows centered on a common row y at columns x_1 in the left image and x_2 in the right image.

2.2 Procedure for computing the CCS

Both images are convolved with a bank of several Gabor filters tuned to different frequencies to equally sample the frequency spectra, and with the corresponding x -partial derivative filter bank. The CCS is computed from the responses in a closed form. The bank of filters is

$$\begin{aligned} c_i(x, y) &= \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} e^{j(u_{0i}x+v_{0i}y)}, \\ c_{xi}(x, y) &= \frac{\partial c_i(x, y)}{\partial x} = \left(-\frac{x}{\sigma^2} + ju_{0i}\right) c_i(x, y), \end{aligned} \quad (2)$$

where $i = 1, \dots, N$ is the index of a filter tuned to a frequency (u_{0i}, v_{0i}) with a constant scale σ .

The input images $f(x, y)$, $g(x, y)$ are convolved with the filters, so the responses are

$$G_{f_i}(x, y) = f * c_i, \quad G_{g_i}(x, y) = g * c_i, \quad G_{xf_i}(x, y) = f * c_{xi}. \quad (3)$$

We estimate the local frequency component, i.e. a partial derivative of the phase of the Gabor response

$$u_{f_i}(x, y) = \frac{\Im(G_{xf_i})\Re(G_{f_i}) - \Re(G_{xf_i})\Im(G_{f_i})}{|G_{f_i}|^2}. \quad (4)$$

Now, the CCS will be evaluated per scanline y . First, for all the cells of the correlation table and for all the filters $i = 1, \dots, N$, we estimate the local subpixel disparity

$$d_i(x_1, x_2) = \frac{\arg(\overline{G_{f_i}(x_1, y)}G_{g_i}(x_2, y))}{u_{f_i}(x_1, y)}. \quad (5)$$

We denote $a_{f_i}(x, y) = |G_{f_i}|$, $a_{g_i}(x, y) = |G_{g_i}|$ and finally, we aggregate the data using the formula:

$$\text{CCS}(x_1, x_2) = \frac{2 \sum_{i=1}^N a_{f_i}(x_1, y) a_{g_i}(x_2, y) e^{jd_i(x_1, x_2)}}{\sum_{i=1}^N a_{f_i}(x_1, y)^2 + \sum_{i=1}^N a_{g_i}(x_2, y)^2}. \quad (6)$$

In this subsection we have described completely the procedure for computing the correlation table of the Complex Correlation Statistic. In the next section we will explain it more in details and show why it works.

2.3 Derivation of the formula for the CCS

Locally, we have two signals $f(x, y)$ and $g(x, y)$ related by

$$g(x, y) = f(x + d(x, y), y), \quad (7)$$

where the (local) disparity is assumed to be small and linearly varying with x, y

$$d(x, y) = d_0 + d_1x + d_2y. \quad (8)$$

First, we will show the $d(x, y)$ can be estimated from the responses of the Gabor filters. Let us assume that our signals are real signals consisting of a single frequency (u, v) only

$$\begin{aligned} f(x, y) &= a \cos(ux + vy + \varphi), \\ g(x, y) &= a \cos(u(x + d_0 + d_1x + d_2y) + vy + \varphi). \end{aligned} \quad (9)$$

Convolving these signals with the Gabor filter tuned to the frequency (u_0, v_0) , we get

$$\begin{aligned} G_f(x, y) &= f * c \approx \frac{1}{2} a e^{-\sigma^2((u_0-u)^2+(v_0-v)^2)} e^{j(u_0x+v_0y+\varphi)}, \\ G_g(x, y) &= g * c \approx \frac{1}{2} a e^{-\sigma^2((u_0-(u+d_1))^2+(v_0-v)^2+2d_2(v-v_0))} e^{j(u(d_0+d_1x+d_2y)+ux+vy+\varphi)}. \end{aligned} \quad (10)$$

This holds well where the signal frequency (u, v) and the tuning frequency of the filter (u_0, v_0) are close to each other. Then we can neglect the contribution of the symmetric frequencies, which arises from the convolution integrals. The disparity is determined from the argument of the Gabor responses:

$$\arg(\overline{G_f}G_g) = -(ux + vy + \varphi) + (u(d_0 + d_1x + d_2y) + ux + vy + \varphi) = u(d_0 + d_1x + d_2y) = u d(x, y), \quad (11)$$

which is the formula (5), because $u = u_f$ is the local frequency, i.e. the derivative of the response phase, as can be seen by taking partial derivatives of the arguments of (10):

$$\begin{aligned} u_f &= \frac{\partial \arg(G_f)}{\partial x} = u, & v_f &= \frac{\partial \arg(G_f)}{\partial y} = v, \\ u_g &= \frac{\partial \arg(G_g)}{\partial x} = u + d_1u, & v_g &= \frac{\partial \arg(G_g)}{\partial x} = v + d_2u. \end{aligned} \quad (12)$$

We can estimate the disparity gradient (d_1, d_2) from the local frequencies as $d_1 = (u_g - u_f)/u_f$, $d_2 = (v_g - v_f)/u_f$. If we use u_g instead of u_f in the denominators here and in (5), we get the complementary disparity \tilde{d} , discussed in Sec. 2.1.

Numerically, we can estimate the local frequency, i.e. the partial derivative of the phase of the response, from the Gabor and Gabor partial derivatives filters. We denote $R = \Re(G_f)$, $I = \Im(G_f)$, $R_x = \Re(G_{xf})$, $I_x = \Im(G_{xf})$. Then

$$u_f = \frac{\partial \arg(G_f)}{\partial x} = \frac{\partial}{\partial x} \arctan\left(\frac{I}{R}\right) = \frac{1}{R^2 + I^2} \left(\frac{\partial I}{\partial x} R - \frac{\partial R}{\partial x} I \right) = \frac{I_x R - R_x I}{R^2 + I^2}, \quad (13)$$

which is the formula (4). Formulas for the other local frequencies can be derived analogously.

Computation of the CCS according to (6) is inspired by the Moravec's normalized cross-correlation [6], which works well for windowed data. We use a similar formula regardless of the fact we work with complex Gabor responses. It is not exactly the same, because there is the necessary frequency normalization (5). After the normalization, the meaning is as follows: Under the correspondence, the local disparity estimates d_i from all the filters should agree and the magnitudes a_{f_i} and a_{g_i} for each filter should also be the same. The formula (6) measures how much this is true.

For example, let us have two images where one is shifted from the other with a constant disparity δ which is smaller than one pixel, $f(x, y)$ and $g(x, y) = f(x + \delta, y)$. No noise is assumed. Then all the local disparities in (5) are $d_i = \delta$ and from (10), we can see that $a_{f_i} = a_{g_i} = a_i$. Substituting this into (6) we obtain

$$\text{CCS}(f(x), f(x + \delta)) = \frac{2 \sum_{i=1}^N a_i a_i e^{j\delta}}{\sum_{i=1}^N a_i^2 + \sum_{i=1}^N a_i^2} = e^{j\delta}. \quad (14)$$

The phase $\arg(\text{CCS}) = \delta$ and the magnitude $|\text{CCS}| = 1$. Clearly, when either the local disparities d_i do not agree or the local magnitudes are different, the $|\text{CCS}| < 1$.

Notice, when the scene is not fronto-parallel, i.e. the disparity is not constant $d_1 \neq 0$, $d_2 \neq 0$, then the magnitudes of the Gabor responses a_{f_i} , a_{g_i} differ in (10). It means the $|\text{CCS}| < 1$, but according to experiments we made, it does not cause serious problems for reasonable slants, see Sec 3.

This derivation is valid only for a mono-frequency signal. But it holds well for general signals containing all frequencies. This is due to high frequency localization of the Gabor filters, proportional to the scale σ . The other reason it works well is that formula (6) aggregates the data from various filter bands and averages out weak response estimates and symmetrically invalid estimates. Therefore, we do not need to use any stability criterion as [4, 19].

2.4 Usage of the CCS and technical notes

For each scanline we get the correlation table of CCS. This is a finite set of the sub-pixel correspondence hypotheses. The task of the matching algorithm is to select a subset. It can be simplified so that, we submit a table of magnitudes $|\text{CCS}|$ to a common discrete algorithm, which establishes the integer matches. Then the sub-pixel disparity is obtained by adding the phase of the CCS to them.

There are high correlation values $|\text{CCS}|$ in the vicinity of the truth (sub-pixel) matches, which are due to the fact that each CCS phase aims at the same point, as in Fig. 1. The maximum $|\text{CCS}|$ is not sharp as a consequence. It might be a problem for some algorithms. So we rearrange the table of magnitudes. The resolution of the table in the direction of the disparity correction is increased twice by adding 1/2 pixel cells. The correlations including the phase are binned in these new cells. Then the magnitude in each bin is determined from the correlation which has the smallest phase. Magnitudes in empty bins are set to zero.

Our filter bank usually contains 50 filters, 10 in the horizontal, 5 in the vertical frequency direction. Gabor filter in the frequency domain is a Gaussian centered at the tuning frequency with a standard deviation proportional to the scale $1/\sigma$. So, the tuning frequencies are selected uniformly from $\pm[0.2\pi, 0.8\pi] \times [0.2\pi, 0.8\pi]$. Too high and too low frequencies are excluded because of aliasing and because of the approximations in (10). The scale of the filters is selected from the range $\sigma \in [2, 5]$, depending on the scene complexity. This is a parameter of the method.

3 Experiments

We will demonstrate the important properties of CCS: the magnitude is discriminable and invariant to image sampling and the subpixel disparity estimation of its phase is accurate.

We made several experiments on both synthetic and real data. We will quantitatively evaluate the performance of CCS and compare to other correlation statistics. Qualitative evaluation as disparity maps of real outdoor scenes will be presented.

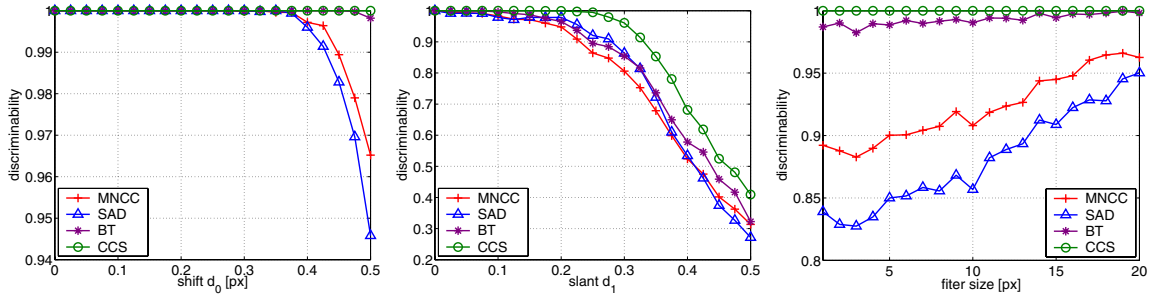


Fig. 3. Discriminability experiment results

3.1 Synthetic Data

We generated stereo-image pairs assuming the cameras observe a textured plane, i.e. the ground-truth disparity is a linear function as in (8). The images were generated in a high resolution, transformed according to the required disparity and finally integrated (filtered) and subsampled into a working resolution.

The high-resolution (1000×1000) image was generated as a random texture containing all frequencies up to the Nyquist limit for the forthcoming subsampling to the working resolution (50×50). This represents a virtual plane in front of the cameras, which was projected into the left image plane with the identity and into the right image plane using the homography $\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \mathbf{h}_3^T]^T$, $\mathbf{h}_1 = [(1 + d_1), d_2, d_0]$, $\mathbf{h}_2 = [0, 1, 0]$, $\mathbf{h}_3 = [0, 0, 1]$ formed according to the required disparity map.

The warped images were convolved with a moving average filter of the same size as the reduction factor ($1000/50 = 20$) and subsampled. This step simulates the integration of the pixel surface in a camera and makes the image spectra more realistic.

In the following experiments, the Gabor scale parameter of CCS was set $\sigma = 2$ and the size of the window of other statistics was set to 5×5 pixels.

Discriminability The discriminability is a property which intuitively says that correlation statistic assigns high values to the true corresponding pairs while keeping low values of all other potential matches.

We define the discriminability as a probability that the ground-truth match has all X-zone competitors [9] of lower correlation, see Fig. 4, and it is estimated from

$$discriminability = \frac{\text{card}\{(i, j) \in \mathcal{G} : \forall (k, l) \in \mathcal{X}(i, j), c(k, l) < c(i, j)\}}{\text{card } \mathcal{G}}, \quad (15)$$

where (i, j) is a cell in the correlation table, \mathcal{G} (red circles) is the set of the ground-truth correspondences and the $\mathcal{X}(i, j)$ (green crosses) is the forbidden zone for (i, j) , and $c(i, j)$ is the correlation value. The estimation was averaged from 100 random trials over texture generation for each stereopair.

This definition of discriminability is insensitive to the scale of the statistic c . We only need the similarity property: higher similarity implies higher value of c .

We will compare the discriminability of CCS with other (window) statistics: the Sum of Absolute Differences (SAD), the Moravec’s Normalised Cross Correlation (MNCC) [6] and the sum of Birchfield-Tomasi sampling insensitive pixel dissimilarities [1] over a window. The SAD and BT is redefined to be a difference of the original statistic from 1 to have the similarity property.

We measured the discriminability for a constant subpixel disparity: $d(x, y) = d_0, d_0 \in [0, 0.5]$ px (fronto-parallel scene), and for a slanted plane: $d(x, y) = d_1 x, d_1 \in [0, 0.5]$. We do not present results for the other slant d_2 , since it was very similar to the behaviour under varying d_1 .

The results for the fronto-parallel case are shown in Fig. 3 (left): The worst case occurs where the true disparity d_0 is 0.5 pixel. The SAD and MNCC tends to fail towards this point, BT has small problems at this point too, although it should be invariant to image discretization, but due to interpolation used in BT, it gets worse as more high frequencies are present, see Fig. 3 (right). We did not observe any failure of CCS in any of the 100 trials, which corroborates its invariance to image sampling.

Unlike the others, the CCS has no problems with high frequencies, see Fig. 3 (right). These plots show the discriminability where $d(x, y) = 0.5$ px versus the size of the moving average integration filter. The size=1 means, there are all frequencies up to the Nyquist limit and they are attenuated with increasing filter size.

The results for the slanted plane are shown in Fig. 3 (center): None of the tested statistics is invariant to the slant. The discriminability decreases with higher slants d_1 as expected, but, surprisingly, the CCS is the best for all measurements.

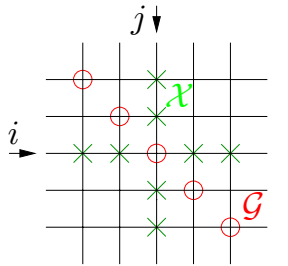


Fig. 4. To the definition of discriminability.

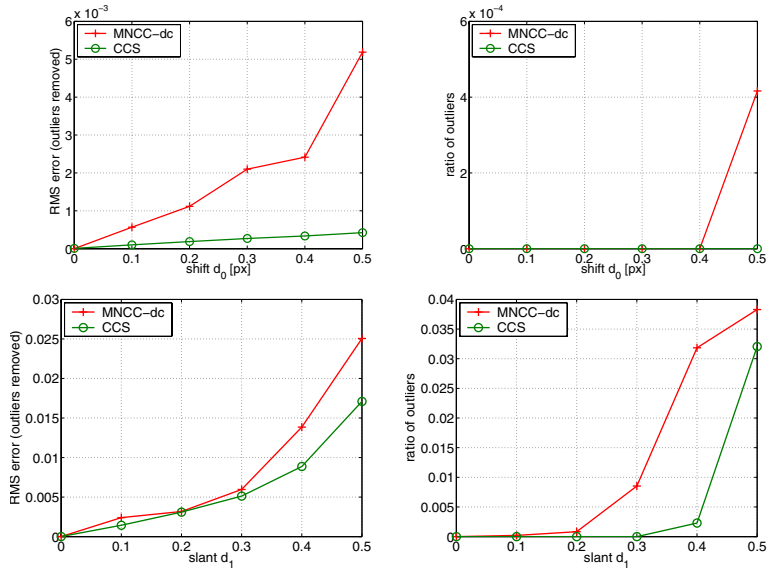


Fig. 5. Accuracy experiment results

Accuracy By matching accuracy we mean that the estimated disparity is close to the ground-truth disparity. We measure the root mean square (RMS) error of this difference, but matches whose error in disparity is higher than 1 pixel are considered outliers and excluded. These outliers are mismatches, a consequence of the low discriminability.

We ran the same integer matching, Confidently Stable Matching (CSM) [9], first with CCS magnitudes, and second with MNCC. We compared the accuracy of the sub-pixel disparity obtained directly from the CCS phase (CCS) with the disparity correction (DC) method on the MNCC matching (MNCC-dc). This method postprocess the integer disparity map by fitting affine window optimizing a ML criterion [8]. Note that RMS results are primarily determined by the DC step, hence we do not need to test other statistics for the pixel resolution matching (like BT, SAD followed by DC, etc.)

We measured the accuracy for the fronto-parallel plane, see Fig. 5 (upper row: RMS error, outlier ratio), and for the slanted plane Fig. 5 (lower row) as in the discriminability experiments. The CCS is clearly the best for the fronto-parallel case and it has no outliers. Surprisingly, the CCS remains better for slanted scenes, despite it does not explicitly model this distortion unlike the DC method.

3.2 Real data

We will report two experiments where the images were captured by real cameras: a precise laboratory test and results on some outdoor scenes. We compared the performance of the CCS with MNCC-dc algorithm as in the synthetic experiments.

Laboratory Test Scene The algorithm performance was tested on the CMP dataset [5]. The scene consists of thin stripes in front of a planar background, see Fig. 7 (first row). The algorithms are evaluated on an image series of varying texture contrast.

We observed the Mismatch Ratio MIR (matches which differ more than one pixel from the ground-truth), the False Negative Ratio FNR (the sparsity of matching) and the accuracy as the RMS error, Fig. 6. We can see the performance of CCS is better for all texture contrasts in all observed quality measurements. It has about twice less mismatches, the FNR is lower by the order of 2 magnitudes (!).² The accuracy is also better: from twice (for weak textures of low contrast, at left) to 1.5 times lower RMS error for high contrasts. The disparity maps for the best texture contrast are shown in fig. 7 (first row).

The whole evaluation method is described in [5], where results for other algorithms are reported.

Outdoor scenes We will show two results from an outdoor scene, which were captured by a standard hand-held digital camera.

²The reason is that for a standard windowed correlation, besides the loss of discriminability where the disparity is around 0.5 pixel offset, there are gaps in the disparity map where the disparity changes, see the unassigned iso-disparity countours in the disparity maps in Fig. 7. The matching algorithm cannot decide which of the two competitive matches having approximately the same correlation to select, and according to the stability principle [9] it decides to reject both of these matches. While in the case of CCS, this undecidable situation does not occur, since all such competitive matches point towards the same match between pixels.

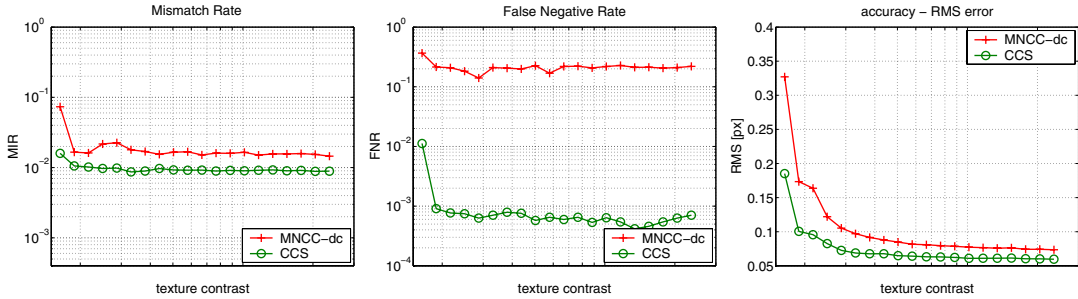


Fig. 6. Laboratory test scene: matching error evaluation plots

The second row of Fig. 7 is a shot from the St. Martin scene with non-planar surfaces, with some half occluded regions and thin objects. The CCS gives denser results than MNCC-dc. The occlusion boundaries are not estimated worse than with a standard correlation. Gabor scale in CCS was set to $\sigma = 4$ and the window size of MNCC was set to 9×9 pixels.

Next example, Fig. 7 (last row), is a scene with an inscription above the door. Importantly, the letter-to-background difference in disparity is less than one pixel. The inscription is clearly visible in the CCS disparity map, while not in the MNCC-dc map, which is very noisy. Gabor scale in CCS was set to $\sigma = 3$ and the window size of MNCC was set to 7×7 pixels.

4 Discussion and Conclusions

We have proposed a new Complex Correlation Statistic which is insensitive to image sampling and provides sub-pixel matching accuracy. We showed on both synthetic and real data it achieves a high discriminability and accuracy of the sub-pixel disparity estimation. In theory, using the CCS, we obtain sampled continuous solution from a discrete algorithm.

The model of the CCS assumes locally fronto-parallel scene (constant disparity) and continuous surface. We showed analytically which distortion occurs when the surface is slanted (locally linear disparity). Practical experiments demonstrated that for this case the distortion is not that strong and we have consistently higher discriminability than standard windowed statistics including windowed Birchfield-Tomasi's dissimilarity. The accuracy is higher even compared to the algorithm which models the distortion of the matching window. On a real data, we showed that it works for curved surfaces as well.

When the surface is not continuous, i.e. when there are occlusion boundaries and thin objects, the algorithm exhibits errors as all other window-based algorithms do: These are the overreach artifacts near occlusion boundaries when the texture of the background is weak [10]. This might be quite a strong effect, although the effective window size due to Gaussian envelope is small, non zero entries in the window may reach far. The scale of the Gabor filter σ should be as small as possible, but large enough to ensure the discriminability and numerical stability of the CCS estimation. Perhaps, the solution might be to design a filter bank which consists of filters with various scales σ . Researchers often choose the σ inversely proportional to the tuning frequency, see e.g. [4, 19], but we have not find a rigorous arguments to do so. Designing an optimal filter bank is a subject of our current research.

The computational complexity of this algorithm is higher than the complexity of standard windowed intensity statistics, because we cannot use the boxing algorithm, which exploits precomputed partial sums. But there are no iterations and optimizations within, we have a closed-form formula for the CCS. So, using a special (but simple) parallel hardware, we can even have a real time implementation.

References

- [1] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *PAMI*, 20(4):401–406, April 1998.
- [2] Gabor D. Theory of communication. *Journal of the IEE*, 93(26):429–439, 1946.
- [3] D. J. Fleet and A. D. Jepson. Stability of phase information. *PAMI*, 15(12):1253–1268, 1993.
- [4] D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, 1991.
- [5] J. Kostková, J. Čech, and R. Šára. Dense stereomatching algorithm performance for view prediction and structure reconstruction. In *Proc. SCIA*, pages 101–107, 2003.

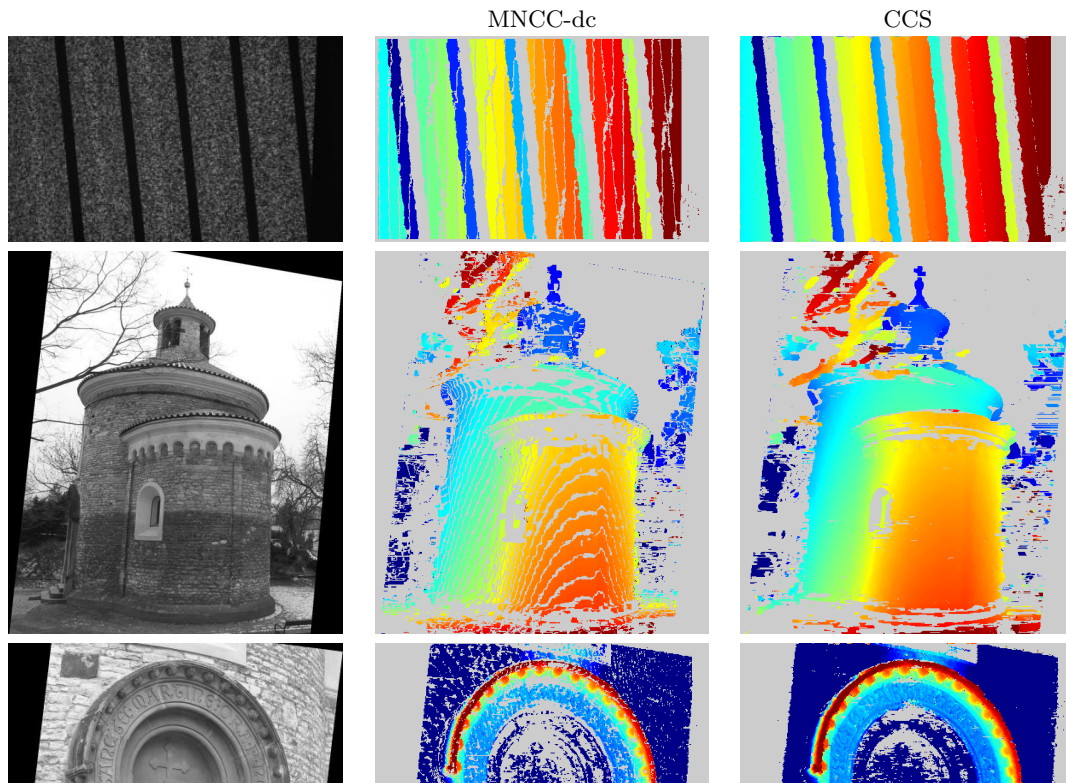


Fig. 7. Real scenes: left images and disparity maps (color-coded, gray are unassigned disparities)

- [6] H. P. Moravec. Towards automatic visual obstacle avoidance. In *Proc. 5th Int. Joint Conf. Artificial Intell.*, pages 584–594, 1977.
- [7] T. D. Sanger. Stereo disparity computation using gabor filters. *Biol. Cybern.*, 58(6):405–418, 1988.
- [8] R. Šára. Sub-pixel disparity correction. Internal working paper 98/01, Center for Machine Perception, Czech Technical University, Prague, 1998.
- [9] R. Šára. Finding the largest unambiguous component of stereo matching. In *Proc. ECCV*, pages 900–914, 2002.
- [10] Radim Šára and Ruzena Bajcsy. On occluding contour artifacts in stereo vision. In *Proc. CVPR*, pages 852–857, 1997.
- [11] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proc. Workshop on Stereo and Multi-Baseline Vision*, pages 131–140, 2001.
- [12] M. Shimizu and M. Okutomi. Precise sub-pixel estimation on area-based matching. In *Proc. ICCV*, pages 90–97, 2001.
- [13] M. Shimizu and M. Okutomi. Significance and attributes of subpixel estimation on area-based matching. *System and Computers in Japan*, 34(12):1791–1800, 2003.
- [14] J. Sun, H.-Y. Shum, and N.-N. Zheng. Stereo matching using belief propagation. In *Proc. ECCV*, pages 510–524, 2002.
- [15] R. Szeliski and D. Scharstein. Sampling the disparity space image. *PAMI*, 25(3):419–425, 2004.
- [16] F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *Proc. ICCV*, pages 900–906, 2003.
- [17] O. Veksler. Fast variable window for stereo correspondence using integral images. In *Proc. CVPR*, pages 556–561, 2003.
- [18] J. Weng. A theory of image matching. In *Proc. ICCV*, pages 200–209, 1990.
- [19] Y. Xiong and S. A. Shafer. Variable window gabor filters and their use in focus and correspondence. Technical Report CMU-RI-TR-94-06, 1994.
- [20] Y. Xiong and S. A. Shafer. Hypergeometric filters for optical flow and affine matching. *IJCV*, 24(2):163–177, 1997.