An Enhanced Correlation-Based Method for Stereo Correspondence with Sub-Pixel Accuracy

Emmanouil Z. Psarakis and Georgios D. Evangelidis

Department of Computer Engineering and Informatics
University of Patras, 26500 Patras, Greece
e-mail: {*psarakis, evagelid*}*@ceid.upatras.gr*

Abstract

The invariance of the similarity measure in photometric distortions as well as its capability in producing subpixel accuracy are two desired and often required features in most stereo vision applications. In this paper we propose a new correlation-based measure which incorporates both mentioned requirements. Specifically, by using an appropriate interpolation scheme in the candidate windows of the matching image, and using the classical zero mean normalized cross correlation function, we introduce a suitable measure. Although the proposed measure is a nonlinear function of the sub-pixel displacement parameter, its maximization results in a closed form solution, resulting in reduced complexity for its use in matching techniques. Application of the proposed measure in a number of benchmark stereo pair images reveals its superiority over existing correlation-based techniques used for sub-pixel accuracy.

1 Introduction

The determination of the 3-D location of objects given different views (images) of the object scene is of great importance in the three dimensional (3-D) object reconstruction problem as well as in a large number of imaging applications. The actual position of a scene element can be determined from the disparity of its two (or more) depicted intensities on the image pair (sequence). However, prior to establishing disparity, the correspondence problem known as the image matching problem must be solved.

Matching techniques according to the strategy they follow to solve the correspondence problem, and thus constructing the disparity map, can be broadly classified into two main categories [3]. More specifically, techniques that construct the disparity map by solving the correspondence problem in a pixel by pixel basis are referred as local, while techniques that consider the correspondence problem as a global optimization problem are referred as global. Among the most well known techniques belonging into the aforementioned categories, are the differential matching, the cross correlation, graph cuts, the global energy optimization method and dynamic programming based methods to name a few.

All local methods, typically use an appropriate measure in order to quantify the existed similarity between the template window and the candidate ones. Widely used similarity measures are the sum of squared differences (SSD), the sum of absolute differences (SAD), and the normalized cross-correlation (NCC) as well as their zero-mean counterparts [7]. Among these measures, only the zero-mean normalized cross correlation is invariant to both shift and scale photometric distortion [7]. This property is required in many stereo vision algorithms, especially in outdoor applications where the illumination of the scene is nonuniform.

Except the invariance in photometric distortions, another feature which is desired in a large number of applications, is the ability of the matching algorithm for producing a disparity map with sub-pixel accuracy. In [2] a comparative study in a registration framework for methods used for sub-pixel accuracy is presented and an iterative scheme for an intensity interpolation method is proposed. The most commonly used approach in stereo matching is based on several polynomial interpolation schemes including correlation interpolation methods, intensity interpolation methods or phase correlation interpolation [1], [4]. However, there exist some limitations resulting from the application of this approach into the correspondence problem [4]. In [6] a similarity measure which is insensitive to image sampling is presented. According to this approach locally linearly inter-

This work is financed by the General Secretariat for Research and Technology of Greek Government as part of the project “XROMA-Handling of non-rigid materials with robots,” PENED 01.

polated templates of the two compared windows are used. A modification of this approach based on interpolated images to produce a disparity map with the desired sub-pixel accuracy is proposed in [5]. However, the computational cost required by the solution of the resulting correspondence problem increases as the interpolation factor increases and thus these methods can not be considered suitable for real-time applications.

In this paper we propose a new similarity measure which is based on the correlation coefficient that we call Enhanced Normalized Cross Correlation (ENCC). More specifically, by using an appropriate linear interpolation scheme on the intensities of two adjacent candidate windows, we succeed in introducing a suitable similarity measure. This measure, although is based on a linear interpolation scheme, does not demand the reconstruction of any intensity value, while at the same time it has infinite precision in its sub-pixel estimates. Although the proposed measure is a non linear function of the sub-pixel displacement its maximization results in a closed form solution.

The paper is organized as follows. In Section 2, we formulate the correspondence problem based on the common framework used in local window-based methods. We also introduce the correlation coefficient and outline some of its most important properties. We define the proposed similarity measure and formulate an appropriate optimization problem for the optimum specification of the desired sub-pixel displacement. We also present with a theorem the optimum solution in closed form. In Section 3, we use the proposed similarity measure in specific image correspondence problems and we compare its performance against the Normalized Cross Correlation (NCC) with parabola fitting technique. Finally Section 4, contains our conclusions.

2 Problem Formulation

Let us consider a rectified stereo image pair of images, with $\mathcal{I}_L(i,j)$ and $\mathcal{I}_R(i,j)$ denoting their intensity functions. Then, the stereo correspondence problem aims at finding a univalued nonnegative disparity map $d_L(i,j)$ such that the following relation approximately holds

$$\mathcal{I}_L(i,j) = \mathcal{I}_R(i,j-d_L(i,j)). \tag{1}$$

In order to solve the image correspondence problem in the framework of a local window-based method, let us consider that $W(n,m)$ denotes an image window of size $N_1 \times N_2$ with its center located at the point with coordinates n,m, and let

$$\mathbf{w}(n,m) = [w_1 \; w_2 \; \dots \; w_{N-1} \; w_N]^t \tag{2}$$

be the vector resulting by stacking up the columns of the window $W(n,m)$, where $N = N_1N_2$ is its length. Let us also define the zero mean normalized version of vector $\mathbf{w}(n,m)$ as

$$\bar{\mathbf{w}}^\circ(n,m) = \frac{\mathbf{w}(n,m) - \bar{w}(n,m)}{\|\mathbf{w}(n,m) - \bar{w}(n,m)\|_2} \tag{3}$$

where $\bar{w}(n,\ m)$ and $\|\mathbf{w}(n,\ m)\|_2$ denote its mean value and Euclidean norm respectively.

By selecting a template window $W_L(n,m)$ in the *reference* image, and a window $W_R(n,m-d)$ in the *matching* image, we can define, using the above notation, their correlation coefficient as the inner product of the vectors

$$\rho_{n,m,d} = \bar{\mathbf{w}}_L^{\circ\mathbf{t}}(n,m)\bar{\mathbf{w}}_R^\circ(n,m-d) \tag{4}$$

and use it as a similarity measure for the centers of the above defined windows.

A remarkable property satisfied by the similarity measure defined in Equ (4) is its invariance to both shift and scale photometric distortions. This property establishes the correlation coefficient as a suitable similarity measure for the image correspondence problem.

Having defined the similarity measure and by assuming that the disparity inside the fixed size window is constant, the solution of the correspondence problem between the center of the windows $W_L(n,m)$ and $W_R(n,m-d)$, results from the solution of the following "winner takes all" [3] maximization problem:

$$\max_{0\leq d\leq R} \rho_{n,m,d} \tag{5}$$

where $\mathcal{R}$ is the disparity range.

By solving the maximization problem for each pixel of the reference image, we obtain a disparity map that contains an estimation of the scene depth in pixel accuracy. Although, in most applications pixel accuracy may be adequate, several problems [8] require sub-pixel accuracy of the disparity map. In the next section we propose a new correlation-based measure capable of producing such increased resolution.

2.1 Sub-Pixel Resolution and the Proposed Measure

We need to redefine the correlation coefficient of Equ (4) in such a way to become a similarity measure capable of producing a disparity map with sub-pixel accuracy. Since we are interested in sub-pixel accuracy, we must reconstruct the candidate windows of the matching image. Usually this goal can be achieved by using some realizable one dimensional interpolation kernel. In such a case it is expected that the correlation coefficient defined in Equ (4) will become a function of the continuous spatial variable τ and the maximization problem solving the sub-pixel correspondence problem will take the following form

$$\max_{0\leq d\leq R}\max_{\tau}\rho_{n,m,d}(\tau). \tag{6}$$

It is clear that the specific form of the correlation function $\rho(\tau)$ and the computational cost of the maximization problem defined in (6), heavily depend on the specific form of the interpolation kernel we use for the reconstruction of the matching windows. Notice also that if the maximization of the correlation function with respect to the displacement parameter τ has a closed form solution, we expect that the computational cost of the total correspondence problem will increase only slightly as compared to the computational cost of the original problem with pixel accuracy. In the opposite case, a maximization algorithm is required resulting in a substantial increase of the computational cost. Alternatively, we can change the order of the two maximizations in (6), and by sampling the continuous spatial variable τ, we can solve the total correspondence problem. Notice though that by following such a strategy, the accuracy of the estimated displacements is bounded while the computational cost increases considerably. To avoid these drawbacks, in the sequel we will incorporate an appropriate first order interpolation kernel into the similarity measure defined in (4). Let us therefore introduce the following N-dimensional vector function

$$\mathbf{w}_R(n,m+\tau)=\mathbf{w}_R(n,m)+\tau(\mathbf{w}_R(n,m)-\mathbf{w}_R(n,m-1)) \tag{7}$$

which is a continuous linear function of the spatial variable τ resulting from the application of a first order interpolation kernel based on the backward differences along each row of the matching windows. Notice that if we spatially sample each element of the N-dimensional vector function $\mathbf{w}_R(n,m+\tau)$ with a sampling rate of M samples per vector element, then we create an $N\times M$ matrix where each column $\tau_n, n=0,1,\ldots,M-1$ constitutes the n-th linearly interpolated (if $\tau_n \ \in[-1\ 0]$) or extrapolated (if $\tau_n \ \notin[-1\ 0]$) sample from adjacent pairs of pixels in the selected windows. In that sense, the vector function defined in Equ (7) can be considered as a linearly interpolated (or extrapolated) function with an infinite interpolation factor. Equ (7) can be rewritten in the following form

$$\begin{aligned}\mathbf{w}_R(n,m+\tau)=&\ \mathbf{w}_R(n,m-1)+(1+\tau)(\mathbf{w}_R(n,m)-\\ &\mathbf{w}_R(n,m-1))\end{aligned} \tag{8}$$

and thus revealing the equivalence of the forward and backward difference operator in the proposed first order interpolation scheme.

In fact, instead of the backward operator, we can use any difference operator $\Phi(\mathbf{w}_R(\cdot\,,m+1),\mathbf{w}_R(\cdot\,,m),\mathbf{w}_R(\cdot\,,m-1))$ which estimates the derivative of the matching windows along its rows. A characteristic example, is the central differences operator.

Our goal now is to incorporate the intensity vector function of τ into the similarity measure defined in Equ (4). To this end, let us define the following correlation function

$$\rho_{n,m,d}(\tau)=\bar{\mathbf{w}}_R^{\circ\mathbf{t}}(n,m-d+\tau)\bar{\mathbf{w}}_L^{\circ}(n,m). \tag{9}$$

Using the definitions of the inner product and zero mean normalized vector of Equ (4), and after some mathematical manipulations, Equ (9) can be rewritten as[1]

$$\rho_d(\tau)=\frac{\rho_d+\tau(\rho_d-\lambda\rho_{d-1})}{\sqrt{(1+\lambda^2-2\lambda r)\tau^2+2(1-\lambda r)\tau+1}} \tag{10}$$

where

$$\lambda=\frac{\|\mathbf{w}_R(n,m-d-1)-\bar{w}(n,m-d-1)\|_2}{\|\mathbf{w}_R(n,m-d)-\bar{w}(n,m-d)\|_2} \tag{11}$$

is the ratio of norms of the adjacent windows and

$$r=\bar{\mathbf{w}}_R^{\circ t}(n,m-d)\bar{\mathbf{w}}_R^{\circ}(n,m-1-d) \tag{12}$$

their correlation coefficient.

Having defined the correlation coefficient as a continuous function of the translation parameter τ, and for a given value d_0 of $d\in[0,\mathcal{R}]$, in the next paragraph we are going to solve the following maximization problem:

$$\max_{\tau}\rho_{d_0}(\tau). \tag{13}$$

2.2 The Optimum Displacement Parameter

Although the correlation coefficient defined in Equ (10) is nonlinear with respect to the displacement parameter τ, its maximization results in a closed form solution. We present the corresponding formula, without proof, in the next theorem.

Theorem 1: Let d_0 be given and λ and r be as they defined in Equs (11) and (12). Suppose that the denominator of Equ (10) is not degenerate then, $\rho_{n,m,d_0}(\tau)$ attains its unique extremum on

$$\tau^0=\frac{\rho_{d_0-1}-r\rho_{d_0}}{\lambda(r\rho_{d_0-1}-\rho_{d_0})+r\ \rho_{d_0}-\rho_{d_0-1}}. \tag{14}$$

The extremum is a maximum, if and only if the denominator of optimum displacement τ° in (14) is negative, and its corresponding value is given by

$$\rho_{d_0}(\tau^0)=\sqrt{\frac{\rho_{d_0}^2+\rho_{d_0-1}^2-2r\rho_{d_0}\rho_{d_0-1}}{1-r^2}}. \tag{15}$$

[1] In order to simplify our notation, from this point on the subscripts $m,\ n$ of the correlation function will be removed.

Using the results of Theorem 1, we can achieve the optimum sub-pixel displacement. We must stress that according to our formulation, the increase in computational cost, as compared to the classical correlation-based method, is negligible.

Having completed the presentation of our similarity measure, in the next section we are going to adapt our results to the image correspondence problem.

3 Simulation Results

In this section we apply the proposed similarity measure in specific image correspondence problems and compare its performance against alternative matching techniques which are based on the classical cross-correlation measure. More precisely, we use the Shimizu-Okutomi image model proposed in [4] in order to evaluate the accuracy of the proposed method and compare its performance, in terms of the resulting displacement estimation error, against the NCC with parabola fitting technique as well as its modification proposed in [4]. Furthermore, using two known sinusoidal models for the generation of artificially translated images, we evaluate the performance of the proposed measure against the NCC measure in a registration framework, in terms of the achieved RMS error for various values of the displacement parameter. Finally, we use the *Map*, the *Sawtooth* and the *Venus* image pairs from the *Middlebury database* in order to evaluate the performance of the proposed measure in stereo matching problems.

A. Shimizu-Okutomi Image Model

The Shimizu-Okutomi image model exclusively depends on the standard deviation σ of a Gaussian distribution function, where $\sigma > 0.7$ (for details see [4]). In this paragraph using this model we are going to evaluate our method in terms of the resulting displacement estimation error against the NCC measure and the modification proposed in [4]. To this end, let us consider that d^o is the optimal solution of the maximization problem defined in Equ (5). Then, in order to evaluate the estimated sub-pixel displacement for the NCC measure using parabola fitting over three consecutive points, we use the following relation [4]

$$\hat{t} = \frac{\rho_{d^o-1} - \rho_{d^o+1}}{2\rho_{d^o-1} - 4\rho_{d^o} + 2\rho_{d^o+1}} \qquad (16)$$

where ρ is the correlation coefficient defined in Equ (4).

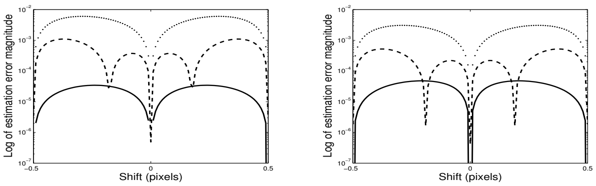

Figure 1. Estimation error of the proposed method (solid line), the parabola fitting method (dotted line) and the Shimizu-Okutomi method (dashed line) for $\sigma = 1.2$ (left) and $\sigma = 1.7$ (right) using the Shimizu-Okutomi image model

Based on this image model, in Figure 1 we plot the error variation in a logarithmic scale for the two values of σ used in [4], as a function of the displacement τ which takes values in the interval $[-.5,\ .5]$. For the computation of the error, we use Equ (14) for the proposed method; Equ (16) for the NCC and its modification proposed by Shimizu and Okutomi, that can cancel out the estimation error by making use of shifted interpolated signal. From the plots it is clear that our approach outperforms the parabola fitting method as well as Simizu-Okutomi modification.

B. Experiments Using Artificial Images

In this experiment we apply two different sinusoidal functions to generate artificially translated images in order to evaluate the RMS estimation error resulting from the application of the proposed measure for various values of the displacement. We also compare the performance of the proposed measure against the NCC measure using parabola fitting. More specifically we use the following sinusoidal forms. *Form I* is a 2-D function proposed in [2] and defined as

$$\begin{aligned} R(i,j) &= 120\frac{\sin(K_x(i-50.1))}{K_x(i-50.1)}\frac{\sin(K_y(j-50.1))}{K_y(j-50.1)} \\ L(i,j) &= R(i-t_i, j-t_j) \end{aligned} \qquad (17)$$

where $K_x = 0.4$ and $K_y = 0.2$. Different values of K_x, K_y cause the images to have different shapes along the respective axes. *Form II* is another 2-D function used in [4] and is given by

$$\begin{aligned} R(i,j) &= \frac{1}{2} + \frac{1}{4}\left(\cos\left(\frac{\pi i^2}{P}\right) + \cos\left(\frac{\pi j^2}{P}\right)\right) \\ L(i,j) &= R(i-t_i, j-t_j) \end{aligned} \qquad (18)$$

where P is the position with spatial frequency equal to 1 [1/pixel]. The value of P is equal to 1000 as in [4]. In our case we do not have translation along the columns, so we consider $t_i = 0$. *Form II* causes heavier distortion between images because of the involution, so we expect an increase in RMS error.

Table 1 contains the RMS error of the methods for different values of t_j. We consider images with size 200×200 while the size of the window is 7×7. We test the algorithms with the shift taking values in the interval $[0,\ 1]$. In Figure 2 the distributions of the estimation of a specific value of t_j obtained by the ENCC and the NCC methods are shown. Notice that the variance of the estimated displacement obtained by ENCC is significantly smaller than the variance of the error obtained by NCC.

Root Mean Square Error				
	Form I		Form II	
shift	NCC	ENCC	NCC	ENCC
0.0613	0.0818	0.0017	0.1145	0.0053
0.1111	0.0800	0.0028	0.1116	0.0088
0.3333	0.0581	0.0064	0.0832	0.0170
0.5000	0.0324	0.0099	0.0590	0.0182
0.8122	0.0758	0.0046	0.1135	0.0122

Table 1. RMS error variation resulted from NCC and ENCC method for various displacements.

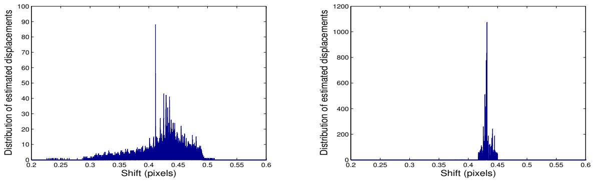

Figure 2. Distribution of estimated displacement obtained by NCC (left) and ENCC (right) for $t_j = 0.4333$ using *Form I.*

C. Experiments Using Stereo Image Pairs

Let us now apply the matching techniques to stereo pairs and use as measure of performance, the percentage of bad matching pixels inside a desired region $\mathcal{R}$ of the reference image. More specifically, following [3] we define

$$B_{\mathcal{R}} = \frac{1}{N_{\mathcal{R}}} \sum_{(x,\ y) \in \mathcal{R}} |d_C(x,\ y) - d_G(x,\ y)| > \delta \qquad (19)$$

where $d_C(x,\ y)$ and $d_G(x,\ y)$ are the computed and the ground true disparity map respectively, δ is the error tolerance and N_R is the number of the pixels belonging in some region $\mathcal{R}$. Assuming that most of the existing local window-based stereo matching algorithms do not produce meaningful results in occluded regions ($\mathcal{O}$) and depth discontinuity regions ($\mathcal{D}$) of the reference image, we exclude the corresponding pixels. We therefore evaluate the measure defined in (19) only inside the regions composed by the intersection of the non-occluded regions and the depth continuous regions, i.e. $\overline{\mathcal{D}} \bigcap \overline{\mathcal{O}}$.

Table 2 contains the bad matching pixel percentage of the corresponding methods for different values of tolerance δ. From Table 1 we observe that the proposed method outperforms NCC for almost all values of the tolerance δ.

Figure 3 depicts the resulting matching errors in disparity maps. Notice the systematic errors exhibited by NCC in some areas of the disparity map, while ENCC gives only spurious outliers which can be easily removed using, for example, a median filter.

Matching Performance — Bad Pixels %						
	Map		Sawtooth		Venus	
δ	ENCC	NCC	ENCC	NCC	ENCC	NCC
0.25	18.34	23.71	27.95	27.46	12.80	16.32
0.50	2.71	3.08	7.97	8.56	3.91	5.05
0.75	1.08	1.22	3.70	4.26	2.75	3.19
1.00	0.74	0.93	1.99	2.49	2.39	2.89

Table 2. Bad matching pixel percentages using $B_{\overline{\mathcal{D}} \bigcap \overline{\mathcal{O}}}$, resulted from the application of the methods under comparison in different stereo image pairs, for different values of the tolerance δ.

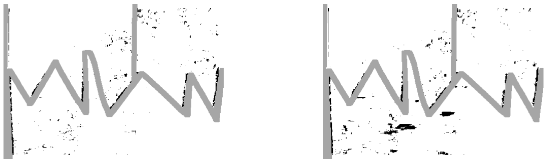

Figure 3. Bad pixels ($\delta = 1$) using ENCC (left) and NCC (right) without sub-pixel refinement (Shaded area contains occlusions and depth discontinuities).

In order to examine the performance of the proposed measure when it is applied to photometrically distorted images, we have nonlinearly distorted one image of the stereo image pairs. In Figure 4, the left image of the sawtooth pair, which has been artificially distorted, and the true disparity map are depicted while Table 3 contains the bad matching pixel percentage. Since we are interested in investigating the behavior of the raw algorithm, we do not take into account any constraints, such as ordering or uniqueness [3] or any post-processing refinement steps. We believe that by using such constraints, the performance of the proposed technique will be substantially improved.

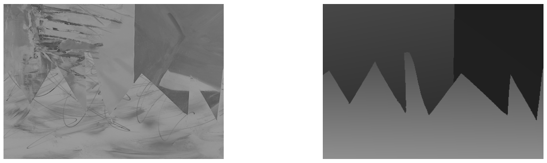

Figure 4. The photometrically adjusted left image and the true disparity map of the *Sawtooth Stereo Image Pair.*

In the last experiment we investigate the behavior of the proposed method in terms of the “pixel-locking” effect. Shimizu and Okutomi define “pixel-locking” as the tendency of the estimated sub-pixel displacements from the parabola fitting technique to concentrate towards integer disparity values. In Figure 5 are shown the distribution of the ground truth disparity values inside the range

Matching Performance — Bad Pixels %						
	Map		Sawtooth		Venus	
δ	ENCC	NCC	ENCC	NCC	ENCC	NCC
0.25	20.18	24.82	29.69	28.58	15.44	18.13
0.50	3.23	3.42	9.40	9.74	5.72	6.54
0.75	1.20	1.35	4.54	4.96	4.32	4.54
1.00	0.82	1.01	2.72	3.07	3.75	4.02

Table 3. Bad matching pixel percentages using $B_{\overline{\mathcal{D}}\bigcap\overline{\mathcal{O}}}$, resulted from the application of the methods under comparison in photometrically distorted images, for different values of the tolerance δ.

$[15-,\ 17+]$ of the sawtooth image pair (available accuracy 0.125 of pixel), and the histograms of the estimated disparities resulting from the application of NCC, Shimizu-Okutomi method and ENCC respectively. From this figure, it is evident that the produced histogram of NCC indeed suffers from the pixel locking effect, while Shimizu-Okutomi method although suppress slightly the undesired "pixel locking" effect, the resulting histogram is far away from the desired ideal one. Our method on the other hand produces an almost uniform histogram which is very close to the distribution of the ground truth disparity values.

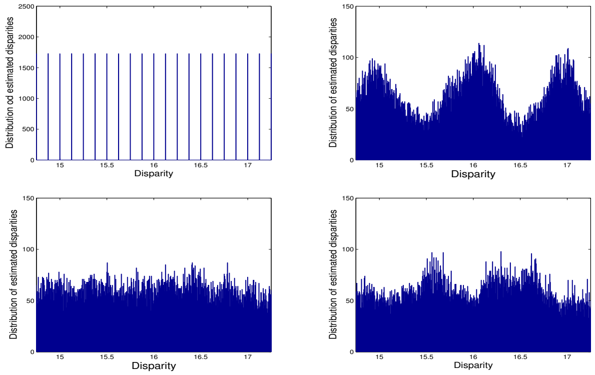

Figure 5. Distribution of ground truth disparity values inside the range $[15-,\ 17+]$ for the Sawtooth Image Pair (up left) and sub-pixel estimated disparities using NCC with parabola fitting (up right), Shimizu-Okutomi method (down right) and ENCC (down left).

4 Conclusions

In this paper we propose a new enhanced correlation-based similarity measure which is invariant in photometric distortions and capable of producing sub pixel accuracy. The optimum value of the displacement results from the solution of a well defined optimization problem and can be computed with the help of a closed form formula. In a large number of simulation examples the proposed method outperforms the well known parabola fitting technique. Further improvement to our scheme can be obtained by incorporating known stereo matching constraints or post-processing refinement steps.

References

[1] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion," *Int'l J. Computer Vision*, vol. 2, no. 3, pp. 283-310, 1989.

[2] Q. Tian and M. N. Huhns "Algorithms for Sub-Pixel Registration," *Comp. Vision, Graphics and Image Procesing*, 35, pp. 220-233, 1986.

[3] D, Sharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision*, vol. 47, no. 1, pp. 7-42, 2002.

[4] M. Shimizu and M. Okutomi, "Presice Sub-Pixel Estimation on Area-Based Matching," *Int'l Conf. Computer Vision*, vol. I, pp. 90-97, 2001.

[5] R. Szeliski and D. Sharstein, "Symmetric Sub-Pixel Stereo Matching," *Proc. European Conf. Computer Vision*, vol. II, pp. 525-540, 2002.

[6] S. Birchfield and C. Tomasi, "A Pixel Dissimilarity Measure that Is Insensitive to Image Sampling," *IEEE Trans. Pattern Analysis and Machine Intelligense*, vol. 20, no. 4, pp. 401-406, Apr. 1998.

[7] O. Faugeras *et al.*, "Real Time Correlation-based Stereo: Algorithm, implementations and applications," *INRIA Technical Report*, No. 2013 August 1993.

[8] S.C. Park, M.K. Park and M.G. Kang "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21-36, May 2003