A New Stereo Method Based on the Detector of Local Structures

Bogusław Cyganek, Jan Borgosz
Department of Electronics and Computer Science
University of Mining and Metallurgy
Al. Mickiewicza 30, Krakow, 30059, Poland
E-mail: {cyganek,borgosz}@uci.agh.edu.pl

Abstract – The paper presents a novel feature-based stereo method that employs a tensor detector of local structures in binary images. The main strength of the presented methods lies in its immunity to false matches as well as very small set of control parameters. Thus, it can be easily implemented on almost every computer platform and be used as a very flexible depth recovery method in navigation or 3D reconstruction.

1 INTRODUCTION

The purpose of this work is to present the novel stereo method that is based on the tensor detector of local structures in binary images. This is a feature-based method in which features at a given scale are reliably given from the tensor detector.

Theoretical background of processing of two images for depth recovery can be found in [2][6][13]. The most troublesome part of the stereo processing is to find corresponding image points. Problems arise mostly due to non-uniqueness of matching points, as well as image noise and distortions, or object occlusions. There are many variations of stereo methods based on choice of primary features for matching. One category, called area-based matching, relies only on intensity comparison of some areas in local vicinity of potentially corresponding pixels. However, much better results are obtained if it is possible to assign certain feature-categories to pixels and then try to match image points. This approach is known as a feature-based stereo method. However, such methods produce only sparse output disparity maps [1][2][6].

In our approach a structural tensor operating in local neighborhoods of binary images was used to detect image features [7][8][4]. As will be shown in this paper, properly converted components of the structural tensor can be used in the further process of stereo matching with very promising results. This happens especially for real stereo images taken by the off-the-shelf cameras.

The paper begins with a brief description of the structural tensor operator for stereo processing. Detailed description of the proposed stereo method follows then. At the end of the paper experimental results have been placed that proved usefulness of the tensor stereo method in real applications.

2 COMPUTATION OF STRUCTURAL TENSOR FOR BINARY STEREO PAIRS

Let us analyze an image with local neighborhood U defined around a point x_0 (Figure 1), where each point has been additionally endowed with a directional vector q, in our case this is an intensity gradient.

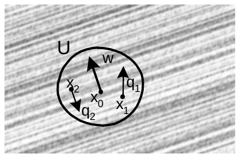


Figure 1. Local neighborhood of a pixel x_0 with shown some gradient vectors q and structural vector w.

The objective is to find such a vector w that fits the best to all other directional vectors q_i from $U(x_0)$. As a measure to compare vectors their inner product is used.

Additionally the following assumptions are defined:

1. An angle and module of w follow signal changes in an image.
2. There is an additional measure of coherency of the local structure.
3. Direction of w is invariant under rotation of π radians.

Based on the above statements, vector w at a point x_0 is an estimator of average orientation in a vicinity $U(x_0)$ that maximizes the following formula [7]:

$$Q = \int_{U(x_0)} \left(q^T(\vec{x}) w(\vec{x}_0) \right)^2 d\vec{x} \tag{1}$$

The square of the inner product in (1) fulfills the invariant assumption on rotation of π radians. Otherwise parallel and anti-parallel configurations of vectors would cancel out.

Let us now introduce a symmetric tensor **T** in the form [4]:

$$\mathbf{T}(\overrightarrow{x_0}) = \int_U \vec{\mathbf{q}}(\vec{x}) \vec{\mathbf{q}}^T(\vec{x}) d\vec{x} \tag{2}$$

where $q(x)\, q^T(x)$ stands for an outer product of vectors, U is a local neighborhood of pixels around the point x_0. Components of **T** can be described by the following formula:

$$T_{ij} = \int_U q_i(\vec{x}) q_j(\vec{x}) d\vec{x} \tag{3}$$

With such a definition of components, the tensor **T** can be put in a matrix form since it is covariant and two dimensional:

$$\mathbf{T} = \begin{bmatrix} T_{xx} & T_{xy} \\ T_{yx} & T_{yy} \end{bmatrix} \tag{4}$$

Taking into account (1) to (4), the problem of finding structural vector w reduces to the solution of the maximization problem, as follows:

$$\max_w(Q) = \max_w(\vec{w}^T \mathbf{T} \vec{w}) \tag{5}$$

Expression (5) is fulfilled for w being an eigenvector corresponding to the maximum eigenvalue of **T**.
It can be shown that the solution of (5) can be described as:

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} T_{xx} - T_{yy} \\ 2T_{xy} \end{bmatrix} \tag{6}$$

Structural vector (6) can be extended by a third component which is equal to the trace of tensor **T**:

$$\mathbf{w}' = \begin{bmatrix} \mathrm{Tr}(\mathbf{T}) \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} T_{xx} + T_{yy} \\ T_{xx} - T_{yy} \\ 2T_{xy} \end{bmatrix} \tag{7}$$

Thanks to this new component in (7) we can distinguish the case where the two eigenvalues λ_1, λ_2 of **T** are equal to zero, i.e. $\lambda_1 = \lambda_2 = 0$ (constant image intensity) from the case $\lambda_1 = \lambda_2 > 0$ (ideal isotropy).

The coherency measure can be defined as a coefficient c [8]:

$$c = \begin{cases} \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 = \frac{\|\mathbf{w}\|^2}{(\mathrm{Tr}(\mathbf{T}))^2} & , \quad \mathrm{Tr}(\mathbf{T}) \neq 0 \\ 0 & , \quad \mathrm{Tr}(\mathbf{T}) = 0 \end{cases} \tag{8}$$

Coefficient c takes on *0* for ideal isotropic structures or for structures with constant intensity value and up to *1* for ideally directional structure. This feature is fundamental for the developed stereo matching algorithm which will be shown further on.

With regard to the stereo processing, it appeared to be easier and more intuitive to use a modified version of structural vector *w'* in the form of vector *s* with components which are defined as follows:

$$\mathbf{s} = \begin{bmatrix} T_{xx} + T_{yy} \\ \angle \mathbf{w} \\ c \end{bmatrix} \tag{9}$$

The first component of (9) is not changed from the corresponding component of *w'* in (7). However, the second component denotes the angle of *w*:

$$\angle w = \begin{cases} \arctan\left(\frac{2T_{xy}}{T_{xx} - T_{yy}}\right) & , \quad T_{xx} \neq T_{yy} \\ \frac{\pi}{2} & , \quad T_{xx} = T_{yy} \quad \wedge \quad T_{xy} \geq 0 \\ -\frac{\pi}{2} & , \quad T_{xx} = T_{yy} \quad \wedge \quad T_{xy} < 0 \end{cases} \tag{10}$$

The third component of (9) constitutes the coherency measure defined in (8).

3 STEREO MATCHING METHOD WITH STRUCTURAL TENSOR

The block diagram of the stereo matching method that utilizes structural tensor detector depicts Figure 2.

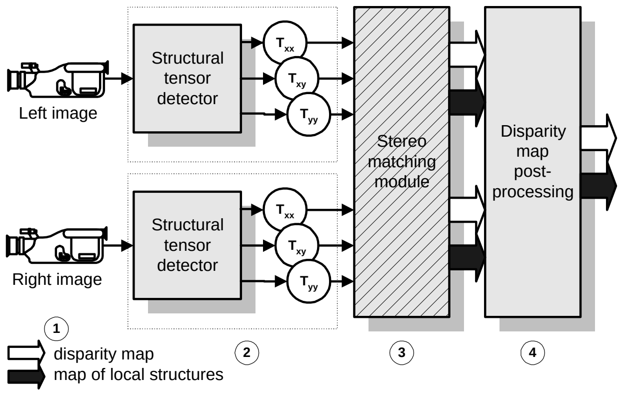


Figure 2. Overall structure of the stereo matching method based on structural tensor detector.

The algorithm begins with stereo pair acquisition modules (block 1 in Figure 2). The canonical (standard) stereo setup is assumed hereafter [2][6]. No camera calibration is assumed *a priori*, although if used for precise 3D measurements this would be a mandatory step. Useful calibration algorithms can be found e.g. in [6][15][1]. Precise photometric calibration is not necessary either, since structural tensor detector can cope with small differences in mean intensity level in both images of a stereo pair [1].

The most distinctive second stage of the algorithm (block 2 in Figure 2) determines local structures in both images separately by means of the structural tensor operator.

Unique features of the output signals from the tensor detector are used during proper stereo matching process, performed by the third module of the algorithm (see Figure 2). Output of this stage constitute two pairs of two maps: disparity map and map of recognized local structures. The pairs differ in reference images, the first one is computed with reference set as a left image of a stereo pair, the second one is set opposite. Such a configuration is led to the last module (block 4 in Figure 2) that is responsible for further processing of output maps, such as disparity map cross-checking [3], or nonlinear filtering [12][11].

An analysis of the coherence component (8) of the structural tensor of many real images reveals that an important binary information could be retrieved by thresholding around median value of an image intensity.

This way we obtain new signal in the form:

$$coh\_bin(I) = B\big(coh(I)\big) \quad (11)$$

where B is a binary thresholding operator around median value, $coh(I)$ denotes the coherence component (9). The binary measure (11) is then processed for further exposition of distinct features, as well as to remove outliers. This is achieved by nonlinear filtering. In our implementation it was a median filter, however it can be easily replaced by e.g. morphological operators. The operation proceeds in accordance with the following equation:

$$coh\_bin\_m(I) = M\big(\mathrm{coh}\_bin(I)\big) \quad (12)$$

where M denotes the nonlinear median filter, whereas $coh\_bin(I)$ is obtained from (11). See also Figure 4.

Based on the properties of the structural tensor operator, the evidence measure for our stereo method has been defined as follows:

$$E(x,y,d) = \begin{cases} \alpha\left|I_l(x,y) - I_r(x+d_x, y+d_y)\right| + \\ +\beta\left|\mathrm{mag}_l(x,y) - \mathrm{mag}_r(x+d_x, y+d_y)\right| + \\ +\gamma\left|\mathrm{angle}_l(x,y) - \mathrm{angle}_r(x+d_x, y+d_y)\right| + \\ +\delta\left|\mathrm{coh}_l(x,y) - \mathrm{coh}_r(x+d_x, y+d_y)\right| \end{cases}, \quad \Psi(x,y,d) \neq 0$$
$$E(x,y,d) = \kappa, \quad \Psi(x,y,d) = 0 \quad (13)$$

where α, β, γ, δ are weight coefficients, $d=[d_x,d_y]$ denotes disparity (for canonical stereo setup $d_y=0$), $I_i(x,y)$ stands for intensity value in the i-th image at point (x,y), $mag_i(x,y)$, $angle_i(x,y)$ and $coh_i(x,y)$ are coefficients of the structural tensor (9), κ denotes a unique value, $\Psi(x,y,d)$ is a binary operator classifying pixels as belonging to an area with local structures (value *1*) and vice versa (value *0*). Value of $\Psi(x,y,d)$ can be determined from the following formula:

$$\Psi(x,y,d) = \mathrm{coh\_bin\_m}_l(x,y) \wedge \mathrm{coh\_bin\_m}_r(x+d_x, y+d_y) \quad (14)$$

where $coh\_bin\_m_i(x,y)$ is value of (12) , whereas $\wedge$ denotes binary *AND*.

The aggregation of evidence measure phase is performed by simple accumulation in a certain areas. This is achieved by simple convolution with a low pass filter mask, such as box or binomial filters [9]. This operation denotes the following equation:

$$\overline{E}(d) = F\big(E(d)\big) \quad (15)$$

where $E(d)$ is the evidence measure (13), F denotes low pass filtering.

Finally, disparity values can be found as a solution to the following minimization problem:

$$D = \min_d \overline{E}(d) \quad (16)$$

where D is a two-dimensional disparity map.

4 EXPERIMENTAL RESULTS

Figure 3 depicts stereo pairs "Trees" and "Pentagon" that were used in experiments.

Figure 3. Stereo pairs "*Trees*" and "*Pentagon*".

Figure 4 presents binarized coherency component processed with the median filter (12) for tested stereo pairs.

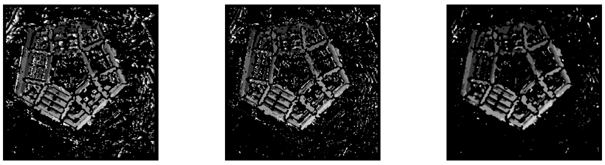
a b

Figure 4. Binarized coherency component processed with nonlinear filter: "*Trees*" (a) and "*Pentagon*" (b).

Figure 5 depicts disparity maps of "*Pentagon*" computed by means of the proposed stereo method. Table 1 contains most important parameters of computation for this stereo pair in our implementation and IBM PC with 800MHz Pentium.

Table 1. Tensor stereo method performance parameters – the "*Pentagon*" stereo pair.

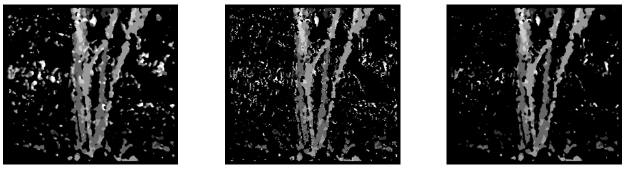
Image res.	Max. disp. range	Structural filter	Aggregation filter	Disp. map post proc.	Exec time [s]
256×256	5	3×3	7×7	median 3×3	2.364



a b c

Figure 5. Right disparity map of "*Pentagon*" (a), cross checked (b), after 3×3 median filter (c).

Figure 6 shows disparity map obtained for "*Trees*" stereo pair. It can be observed that the cross checked map processed further with the 3×3 median filter does contain only minor distortions what is sufficient for most applications.



a b c

Figure 6. Stereo pair "*Trees*": left disparity map processed with 3×3 median filter (a), cross checked disparity map (b), cross checked disparity map processed with 3×3 median filter.

Table 2 contains performance parameters of the presented method for "*Trees*" stereo pair.

Table 2. Tensor stereo method performance parameters – the "*Trees*" stereo pair.

Image res.	Max. disp. range	Structural filter	Aggregation filter	Disp. map post proc.	Exec time [s]
256×233	5	5×5	9×9	none	2.21

The maximum disparity range parameter is actually the only one that must be specified to the algorithm for its proper operation. Its value depends on supplied images. In our implementation a heuristic method [1] based on variogram analysis [14] has been applied for automatic choice of this parameter.

5 CONCLUSION

The novel stereo method based on tensor detector of local structures presents new alternative to the fast computation of a disparity map for real images. It has been checked empirically that the method performs well for images taken with the off-the-shelf cameras, while its execution time is not very extensive. Especially the version of the method with cross checking and disparity map post processing by means of a nonlinear filter is very promising since the output disparity map does not contain many distortions and false matches. The method has been checked for many real stereo pairs and compared to the classical Marr-Poggio-Grimson [5] and Shirai [10] feature based stereo algorithms. In all cases it performed better. It should be also emphasized that the presented method for proper operation needs only one input parameter to be set properly (i.e. maximum disparity range).

ACKNOWLEDGEMENT

This work has been sponsored by the Polish National Research Committee grant entitled "Correlation of stereo images with tensor operator of local structures".

REFERENCES

[1] B. Cyganek, Stereo matching with tensor representation of local neighborhood, Ph.D. thesis, University of Mining and Metallurgy, Krakow, Poland, 2001.

[2] O. Faugeras, Three-dimensional computer vision. A geometric viewpoint, MIT Press, 1993.

[3] P.Fua: "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features," INRIA Technical Report No 1369, 1991.

[4] G.H. Granlund, H. Knutsson, Signal Processing for Computer Vision, Kluwer Academic Publishers, 1996.

[5] L.L. Grewe, A.C. Kak, Stereo Vision. Handbook of Pattern Recognition and Image Processing: Computer Vision. Edited by T.Y. Young. Volume 2, Academic Press, 1994.

[6] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.

[7] H. Haußecker, B. Jähne, "A Tensor Approach for Local Structure Analysis in Multi-Dimensional Images". Interdisciplinary Center for Scientific Computing, University of Heidelberg, 1998.

[8] B. Jähne, Digital Image Processing. 4th edition, Springer-Verlag, 1997.

[9] B. Jähne, Practical Handbook on Image Processing for Scientific Applications, CRC Press, 1997.

[10] R. Klette, K. Schlüns, A. Koschan, Computer Vision. Three-Dimensional Data from Images. Springer-Verlag, 1998.

[11] R. Klette, P. Zamperoni, Handbook of Image Processing Operators. John Wiley & Sons, 1996.

[12] S.K. Mitra, G.L. Sicuranza, Nonlinear Image Processing. Academic Press, 2000.

[13] D. Scharstein, View Synthesis Using Stereo Vision. Lecture Notes in Computer Science 1582, Springer-Verlag, 1999.

[14] J.-L. Starck, F. Murtagh, A. Bijaoui, Image Processing and Data Analysis. The Multiscale Approach. Cambridge University Press, 2000.

[15] E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision. Prentice-Hall, 1998.