

Depth Recovery with an Area Based Version of the Stereo Matching Method with Scale-Space Tensor Representation of Local Neighborhoods

Bogusław Cyganek

University of Science and Technology
Department of Electronics
Al. Mickiewicza 30, 30-059 Kraków, Poland
cyganek@uci.agh.edu.pl

Abstract. Depth recovery is one of the classical problems of computer vision. Many methods have been developed that address this issue. However, each of them has specific behavior depending on the image contents, acquisition system, computation time, etc. The variety of methods is due to the non uniqueness of the matching process involved in depth recovery from a pair of images of a given scene. The presented method relies on tensor representation of local structures in images. This representation allows for extraction of the local phase, signal intensity and coherence measure of a local neighborhood in different scales. This operator showed to be very useful in detection of locally oriented structures and corners in single digital images – presented at the previous ICCS. In this paper, the tensor operator is used for area-based matching of stereo images. Due to tensor transformation of the input images the more reliable matching was achieved. The method was tested with many different classes of stereo images. In this paper we present and discuss the experimental results and also details of implementation.

1 Introduction

The presented stereo method is an area-based version (dense disparity maps) of the feature-based method presented in [4]. The method relies on tensor representation of local structures in the input images. The tensor operator was already used for detection of local structures and corners in *single* images [3]. Such tensor representation provides the local phase, signal intensity and coherence measure of a local neighborhood in different scales. Thus, this representation allows for more reliable matching than other simple area-based methods. Furthermore, with the tensor method the first step of segmentation into feature and featureless areas is obtained automatically. In effect, based on at least two images of a scene, the method makes possible segmentation of the nearest objects from the background. The presented method was tested in the navigation system for the vision impaired persons [5]. It allowed for direct localization of the nearest objects from the observed scene. The implementation with the novel winner-update matching [2] allowed for speed improvement of the version from [5].

2 Tensor Detector of Local Structures in Pixel Neighborhoods

A detailed derivation of the tensor detector of local structures, as well as its practical implementation, were already published at the previous ICCS [3]. In this paper then we present only the final formulas. The detector of local structures in images takes the following vector form:

$$\mathbf{s} = \begin{bmatrix} T_{xx} + T_{yy} \\ \angle \mathbf{W} \\ c \end{bmatrix}, \quad \text{where} \quad \angle \mathbf{W} = \begin{cases} \arctan\left(\frac{2T_{xy}}{T_{xx} - T_{yy}}\right), & T_{xx} \neq T_{yy} \\ \frac{\pi}{2}, & T_{xx} = T_{yy} \wedge T_{xy} \geq 0 \\ -\frac{\pi}{2}, & T_{xx} = T_{yy} \wedge T_{xy} < 0 \end{cases} \quad (1)$$

The third component c of (1) constitutes a coherency measure [10]:

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

The structural tensor \mathbf{T} , its eigenvalues $\lambda_{1,2}$, as well as its components T_{xx} , T_{xy} , and T_{yy} are defined in [3].

3 Dense Stereo with the Fast Block Matching

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

$$E(x, y, d) = \begin{cases} \alpha \left| I_l(x, y) - I_r(x + d_x, y + d_y) \right| + \\ + \beta \left| s_{1l}(x, y) - s_{1r}(x + d_x, y + d_y) \right| + & \Psi(x, y, d) \neq 0 \\ + \gamma \left| s_{2l}(x, y) - s_{2r}(x + d_x, y + d_y) \right| & \\ \kappa & \Psi(x, y, d) = 0 \end{cases} \quad (3)$$

where α, β, γ 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, $\Psi(x, y, d)$ is a binary operator classifying pixels as belonging to an area with local structures (value 1) and vice versa (value 0); κ is a special value to denote those areas in images. Details of the function $\Psi(x, y, d)$ are provided in [4].

The canonical setup was assumed, thus the epipolar lines were known beforehand. Averaging of the evidence measure was done with a moving average filter in a rectangular window – usually a square from 3×3 to 17×17 pixels wide.

4 Experimental Results

The method was implemented in C++ and all the experiments were performed on the PC with Pentium 1.8GHz. The winner-update strategy [2] was used for matching.

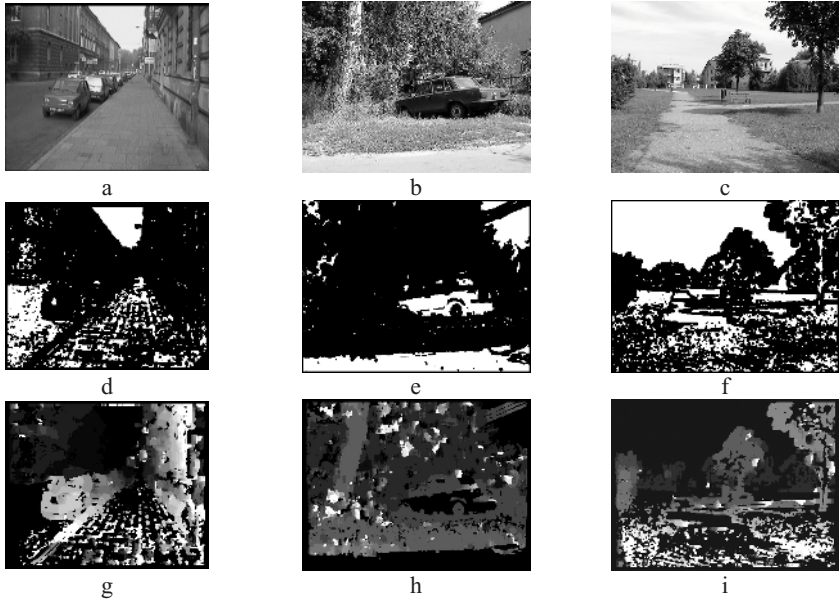


Fig. 1. Stereo matching of real images: the original left images (a,b,c), regions classified as textureless (in white) (d,e,f), disparity maps (g,h,i)

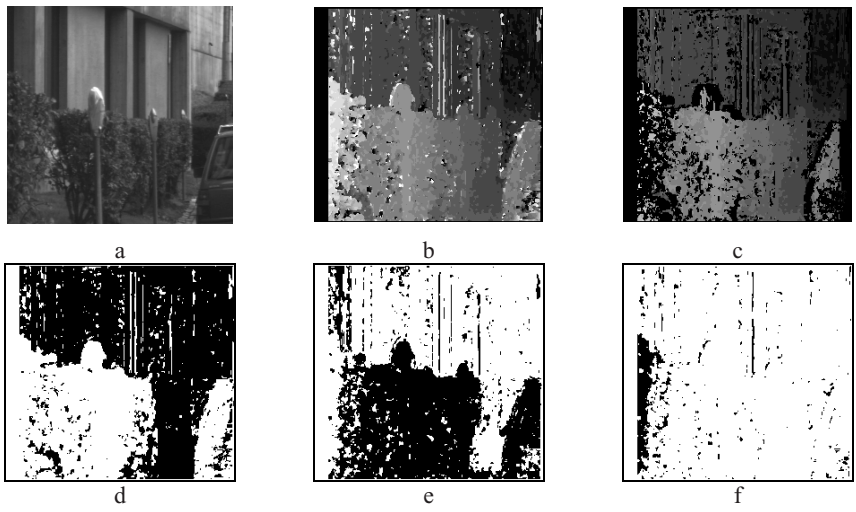


Fig. 2. Depth recovery and segmentation: left image (a), the disparity map with 3×3 matching windows (b), the disparity map after cross-checking (c), segmentation from disparity – background objects (in black) (d), middle objects (e), the closest objects (f)

Fig. 1(a,b,c) presents three left images of the real stereo pairs. The detected textureless areas – in white – are presented in Fig. 1(d,e,f). This level of segmentation is essential to obtain the disparity maps with a good quality – Fig. 1(g,h,i).

Fig. 2 presents experiments with the “Parkmeter” (a). The two disparity maps, obtained with the 3×3 matching windows, are presented in Fig. 2b and Fig. 2c. The latter is validated by a cross-checking. Fig 2.(d,e,f) contain objects (in black) detected by segmentation: background objects (d), middle range objects (e), the closest (f). Original image size is 256×240 pixels. Execution times for disparity maps are as follows: 1.54 s for Fig. 2b, and 3.18 for Fig. 2c.

5 Conclusions

In this paper the depth recovery method from multiple views of the same scene was presented. The dense disparity map was computed by the stereo method that relies on the tensor representation of local neighborhoods of pixels. This representation has the plausible feature of partitioning the input images into areas with sufficiently strong structures, at the same time rejecting featureless areas that cannot be reliably matched. Stereo matching is performed in the domain of the structural tensors by means of the L_1 measure which is computationally efficient. Computation time was also reduced at this stage by application of the winner-update technique for matching. The experimental results performed with different real images showed their robustness, especially in avoiding textureless areas.

References

1. Brown, M.Z., Burschka D., Hager G.D.: Advances in Computational Stereo. IEEE Trans. PAMI, Vol.25, No. 8, (2003) 993-1008
2. Chen, Y-S., Hung, Y-P., Fuh, C-S.: Fast Block Matching Algorithm Based on the Winner-Update Strategy. IEEE Trans. On Image Processing, Vol. 10, No. 8, (2001) 1212-1222
3. Cyganek, B.: Combined Detector of Locally-Oriented Structures and Corners in Images Based on a Scale-Space Tensor Representation of Local Neighborhoods of Pixels, in LNCS 2658, Proceedings of the ICCS 2003 (2003) 721-730
4. Cyganek, B.: Novel Stereo Matching Method That Employs Tensor Representation of Local Neighborhood in Images, Machine Graphics & Vision, 10/3 (2001) 289-316
5. Cyganek, B., Borgosz, J.: Computer Platform for Transformation of Visual Information into Sound Sensations for Vision Impaired Persons, in LNCS 2626 (2003) 182–191
6. Egnal, G., Wildes R.P.: Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches. IEEE Trans. PAMI, Vol. 24, No. 8, (2002) 1127-1132
7. Fua P.: A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features, INRIA Technical Report No 1369 (1991)
8. Hartley, R.I., Zisserman A.: Multiple View Geometry in Computer Vision. CUP (2000)
9. Haußecker, H., Jähne, B.: A Tensor Approach for Local Structure Analysis in Multi-Dimensional Images. Interdisciplinary Center for Scientific Computing, University of Heidelberg, (1998)
10. Jähne, B.: Digital Image Processing. 4th edition, Springer-Verlag, (1997)