PROFESSIONAL STUDIES: Theory and Practice
2016 / 1(16)

COMPARISON OF STEREO VISION ALGORITHMS

Alexander Abramovich
Šiauliai University
Lithuania

Donatas Dervinis
Šiauliai State College, Šiauliai University
Lithuania

Annotation

The aim of the paper is to analyze and compare the stereo vision algorithms by the proposed methodology. The experimental results objective shows the evaluation of 7 different stereo vision algorithms.

Key words: *image processing, image databases, machine vision, stereo vision.*

Introduction

The field of robotics is receiving a great deal of attention. The general robotics area is computer vision or machine vision - hardware and software solutions available, allowing to obtain a digital image of it to process and present the results in a format that can be practically used for real-time machines[1]. Computer vision is divided into monocular and binocular. They respectively are processing flat and three-dimensional images. Stereo vision is popular in different projects: recreational, scientific or industrial. A three-dimensional image is created by technical – a video camera, recording equipment and software tools, that from two the differential images create a stereo view. New approaches are presented every year. The stereo vision methods many authors use by default tests the effectiveness by counting error pixels rates and program running time [2, 9].

The aim of this work is to describe, test and compare several different stereo vision algorithms for three-dimensional image and the evaluation method developed by the evaluation of the supporting video and isolated pixels located at different depths matching.

Spatial stereo vision models

By default a stereo system is designed from two video cameras, which focus to the same area. The standard model is shown in 1 Figure.

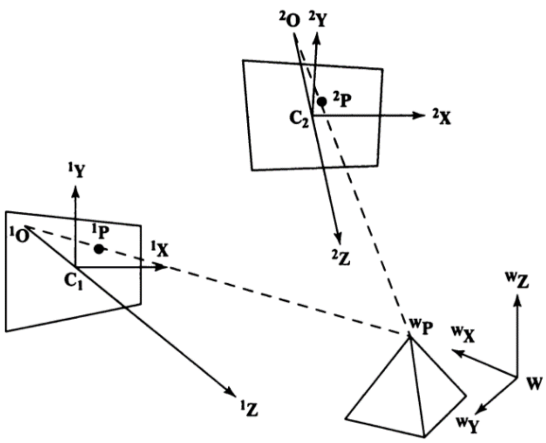


Fig. 1. Standard computer model of the stereo vision

Video cameras coordinate system C1 and C2. is set, that axis Z is minimum distance between object and camera. For calculating 3D coordinates of the object we need to:

– know video cameras C1 and C2 coordinates on W coordinates system and focal length;
– determine ^{W}P compliance with images ^{1}P ir ^{2}P;
– calculate ^{W}P from $^{W}P^{1}O$ ir $^{W}P^{2}O$.

Points ^{1}P ir ^{2}P (Fig. 1) are projected on video cameras matrixes. One camera is called as *reference* (camera projection of matrix shown as π_R in Figure 2), other – as *target* (projection of matrix shown as π_T in Fig. 2). We need to find in both cameras projections these same points of the object P.

Two points Q and P are in the same line (2 Fig. red dotted lines). On reference camera Q and P are same i.e. in π_R we see only one point p ≡ q. In camera matrix π_T – this same point is separate as q ir p.

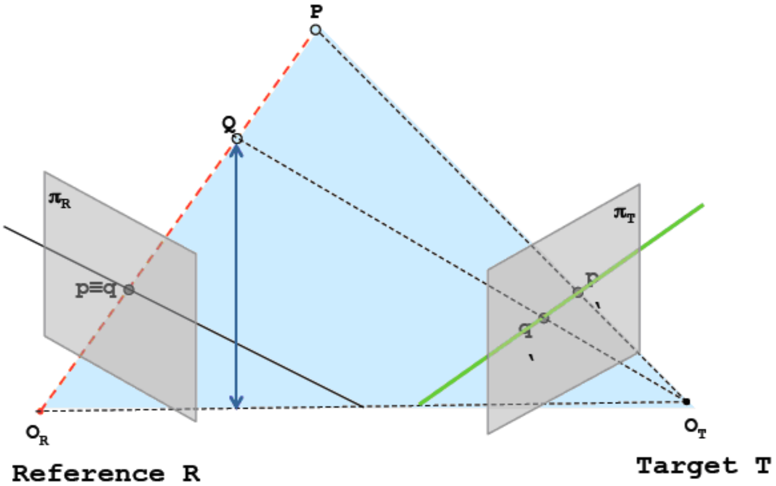


Fig. 2. Three-dimensional model of non-calibrated camera plane

For both camera (i.e. perspectives areas π_T and π_R) are making stereo calibration – standardization. After standardization area π_T and π_R is coincide and stereo system is transformed to one-dimensional system (Fig. 3). Usually transformations are virtually.

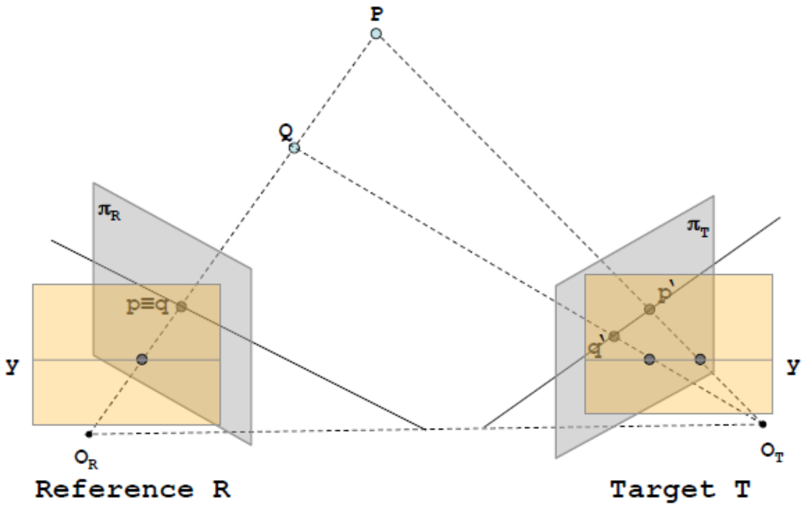


Fig. 3. Standardized model of stereo vision system

Disparity – image depth, other parameter which must be calculate after standardization. Standard method is based triangulation – is calculated triangles PQ_RQ_T and Ppp‘ (Fig. 3).

$$\frac{B}{Z} = \frac{(b + x_T) - x_R}{Z - f}; \tag{1}$$

$$Z = \frac{b \cdot f}{x_R - x_T} = \frac{b \cdot f}{d}, \tag{2}$$

here:
b – basic axis, *f* – focal length; $x_R - x_T = d$ –difference of coordinate of wanted point(P)

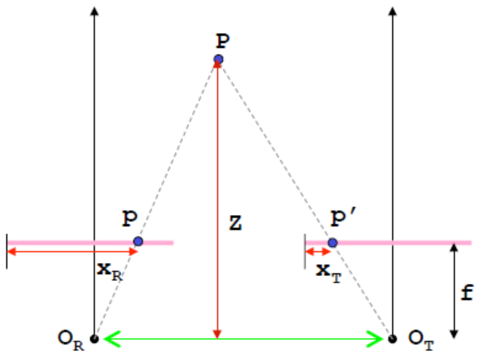


Fig. 4. Model of image depth calculation

The disparity refers to the distance between two corresponding points in the left and right image of a stereo pair. Often as result disparity is grayscale image (Fig. 5 c). Images databases which is compiled to the comparison algorithms, have the synthetic image – *ground-truth* image [4, 9]. If ground-truth grayscale (Fig. 5 d) is whiter – the object (point) is closer and disparity bigger, if darker – object is forward and disparity – lower.

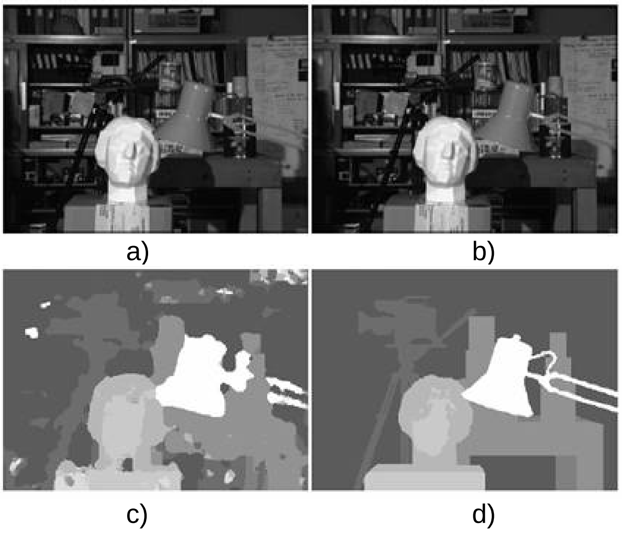


Fig 5. Tsukuba stereo image pair: a – left image, b – right image; c – calculated disparity; d – ground-truth [9]

The horopter is the range of depth values within which objects can be measured by the stereo vision system. The horopter is defined by minimum and maximum disparity values. All objects (or points) are possible to arrange in range [D_{min}, D_{max}] for this is need have a basic axis and focal length (Fig. 6 red – D_{min} and blue – D_{max} dotted line). Available horopter can be discretized e.g. divide to parallel flatness. For better discretization result is possible to use sub-pixel estimation [6]. Sub-pixel estimated horopter is shown as solid lines in Figure 6.

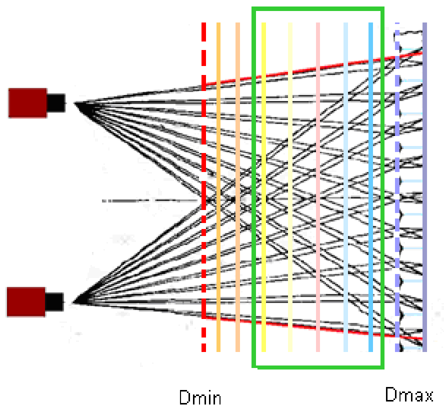


Fig. 6. The horopter, here dotted red and dotted blue line – limits of horopter; green rectangles – horopter area; solid lines between D_{min} and D_{max} – sub- pixel discetisation.

Standard stereo image calculation steps:

1. disparity refinement matching cost computation;
2. cost (support) aggregation;
3. disparity computation and optimization;
4. disparity refinement

The local window based algorithm (or *Pixel-based)* uses 1-3 steps, if worked with strategy "winner takes all". A local based algorithm calculates differences of point and results belong only from points value [3, 5]. Gaussian, Wiener and other filters are used to increase accuracy and quality. These add increased signal to noise ratio. Limitations of this algorithm may correspond to several points in *reference with* the same parameters in a *target* image [3].

Pixel-based matching costs algorithm support regions, often referred to as support or aggregating windows, which could be square or rectangular, fix-sized or adaptive ones. The aggregation of the aforementioned cost functions, leads to the core of most of the stereo vision methods, which can be mathematically expressed as follows, for the case of the sum of absolute differences (SAD):

$$C(x,y,d) = \sum \left| I_R(x,y) - I_T(x+d,y) \right| . \tag{3}$$

– Sum of Square differences (*SSAD)*:

$$C(x,y,d) = \sum \left(I_R(x,y) - I_T(x+d,y) \right)^2 . \tag{4}$$

– Sum of truncated absolute differences (*STAD):*

$$C(x,y,d) = \sum \min \left\{ \left| I_R(x,y) - I_T(x+d,y) \right|, T \right\} . \tag{5}$$

Other global algorithms [8] make explicit smoothness assumptions and solved optimization tasks. Global used only 1, 3, 4, and don't used aggregation step. They are generally based on energy minimization methods. The goal is to find a disparity function *d* which minimizes the "global" energy:

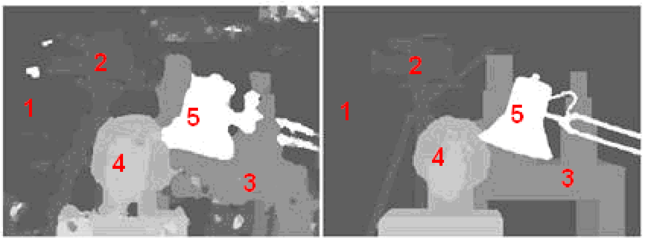
$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) , \tag{6}$$

here $E_{data}(d)$ function show how well disparity function corresponds to a stereo pair. $E_{smooth}(d)$ shows the function selected smoothness filtering algorithm level.

The images which are used in work are created by Tsukuba University – synthetic "tsukuba", on image in different depths is shown: head, lamp and table. This image is synthetic. It is created with special spatial software and for this reason is possible generate and get different versions of the image including the view angle, lighting, size of object and other. One of version is shown in figure 5 a) – left, b) – right images.

Proposed comparison methodology

In figure 7 we have 5 different objects in different depth levels. The lamp is closer; the camera and background is forward. Objects also overlap one another, therefore it's very important after stereo recognition separate object boundaries. On this paper is used methods (described above) – Local methods: standard Basic approach and Basic approach with sub-pixel estimation are both realized with mathematical models SAD, SSD, STAD; Global method is based on Dynamic Programming [8]



a) b)

Fig. 7. Detected objects which are checked (1 –Background, 2 – Camera, 3 – Table, 4 – Head, 5 – Lamp). Calculated image (a) and ground-truth image (b).

Proposed comparison algorithm:

1. On ground–truth image is a detected pixel which belongs to objects (pixel is detected automatic by edge detection with manual corrections). Object pixels is set as $B_n(x,y)$, where n – object number, $n \in [1-5]$.

2. The original image pair is calculated by method (on this paper is used 7 methods – algorithm). As result is getting disparity of stereo images (Fig. 7 a).

3. For each basic approach method the algorithm is finding the optimal window size and the disparity level which is used for further calculation. For this is calculating correlation C between calculated disparity image D and ground-truth all image G:

$$C(s,d) = corr(D(s,d),G); \quad (7)$$

There s – windows size is increasing in step 2 (until 15 pixel), d – disparity in range from 6 to 19 (6 is minimal disparity level for the examined image). As result in 8 figure is shown basic approach SAD algorithm. For the next step, select results that have the highest correlation, lowest window size and lowest disparity (Fig. 8 result with white line).

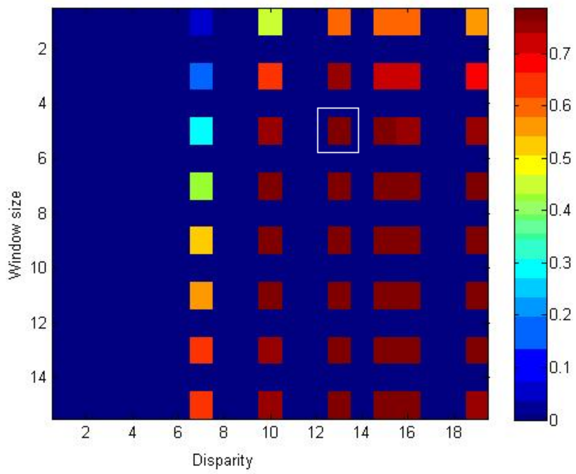


Fig. 8. Stereo vision algorithm basic approach SAD calculated image and ground–truth image correlation of all image versus windows size and disparity level. Result with white line – better, with optimal windows size and disparity.

4. On calculating images find the same boundaries as in the ground–truth image (step 1). The objects pixels is set as $O_n(x,y)$, where n is the object number, $n \in [1-5]$. In figure 7, numbers are shown in the same place on both images: the calculated image (a) and the ground-truth image

5. The correlation between calculation, ground–truth objects $B_n(x,y)$ and calculated image objects with best result $O_n(x,y)$:

$$C_n = corr(B_n(x,y), O_n(x,y)). \quad (8)$$

The algorithm was released on software MatLab2013b, without special Toolbox. Computer specification home style laptop with CPU AMD A8-4500M, RAM 4 Gb, OS Windows 7. Time Pictures size – 288 x 384 pixels.

Result of comparison of 7 stereo algorithms is shown on table 1. As shown better result is when basic approach windows size 3 – 5 and initialized disparity level 13 – 15, approximately execution time is same 36-37 s. Best result getting when used SAD and SSAD algorithm (all image coefficient is greater 0.80). The STAT algorithm windows size can be smaller than SAD or SSA algorithm. By this methodology it is clear that sub-pixel algorithm execution time is longer 15 – 30 % as normal basic approach algorithms, but quality of disparity is approximately same. Advantage of sub-pixel method is not very great.

Algorithm based dynamic programming is slower 50 – 80% versus basic approach. The detection result is better with closer objects (head and lamp) and

The background of image the most difficult to define (boundaries is not clear) and for this reason the result is poor (less 0.25) and unstable.

Table 1

Comparison of stereo vision algorithms

Algorithms	Window size	Dis-parity level	Execu-tion time, s	The Correlation Coefficient					
				All image	Back-ground	Camera	Table	Head	Lamp
Basic aproch SAD	5	13	38.5	0.81	0.22	0.61	0.79	0.58	0.57
Basic aproch SSAD	5	13	36.2	0.80	0.18	0.62	0.79	0.55	0.57
Basic aproch STAD	3	15	36.2	0.76	0.03	0.32	0.77	0.56	0.52
Sub-pixel SAD	7	15	52.8	0.79	0.25	0.62	0.81	0.54	0.58
Sub-pixel SSAD	7	15	48.9	0.78	0.24	0.67	0.80	0.50	0.57
Sub-pixel STAD	3	15	40.6	0.76	0.03	0.36	0.77	0.55	0.53
Dinamic programming	ns	13	59.2	0.82	0.13	0.60	0.77	0.65	0.59

Conclusions

The stereo correspondence problem remains an active area for research. It is therefore expected, that this work will complement the stereo image comparison methods.

The released methodology can be applied to compare various stereo algorithms. The experimental results objective shows an evaluation of 7 different algorithms: 3 locals, 3 locals with sub-pixel estimation and global.

This method allows estimating the program execution time, detection accuracy (different object depth), was setting optimal windows size and disparity levels for the standard algorithm. It reasons that the program was implemented using MATLAB programming without special Toolboxes – execution times are long, but using e.g. special Computer vision hardware (FPGA) and software solutions (OpenCV), the execution time proportionally decrease.

References

1. Lazaros, N., Sirakoulis, G. C., Gasteratos, A. (2008). Review of stereo vision algorithms: from software to hardware. *International Journal of Optomechatronics*, 435–462.

2. Scharstein, D., Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal Computer Vision*, Vol. 47 (Issue 1-3), 7–42

3. Mattoccia, S. (2009). A locally global approach to stereo correspondence. *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*.

4. Martull, S., Peris, M., Fukui, K. (2012). Realistic CG Stereo Image Dataset with Ground Truth Disparity Maps. *ICPR2012 workshop TrakMark*, 40-42.

5. Tombari, F., Mattoccia, S., Stefano, L. Di. (2007). Segmentation-based adaptive support for accurate stereo correspondence. *IEEE Pacific-Rim Symposium on Image and Video Technology PSIVT*, Santiago, Chile.

6. Morgan, G. L., Liu, J. G., Yan, H. (2010). Precise subpixel disparity measurement from very narrow baseline stereo. *IEEE Trans. Geoscience and Remote Sensing*, 48(9), 3424-3433.

7. Psarakis, E. Z. Evangelidis, G. D. (2005). An enhanced correlation-based method for stereo correspondence with subpixel accuracy. *IEEE Int. Conf. on Computer Vision (ICCV'05)*.

8. Criminisi, A., Blake, A., Rother, C., Shotton, J., Torr, P. H. S. (2007). Efficient Dense Stereo with Occlusions for New View-Synthesis by Four-State Dynamic Programming. *International Journal of Computer Vision*, Vol. 7 1, Issue 1, 89-110.

9. Scharstein, D., Szeliski, R. *The Middlebury Computer Vision Pages. Image database*. Link: http:// http://vision.middlebury.edu/..

Received: 23 December 2015
Accepted: 29 February 2016