

# Combined method for calculating the disparity value on stereo images in problems of stereo-range metering

A.N. Volkovich<sup>1</sup>

<sup>1</sup>United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Surganova str. 6, 220012, Minsk, Republic of Belarus

---

## Abstract

The paper considers the solution of the problem of restoring three-dimensional information based on stereo images. An original combined approach to disparity calculation is proposed, as well as a variant of solving the problem of the heterogeneity of the initial data in calculating the actual metric parameters.

*Key words:* stereo images; disparity maps; gradient operators; satellite photographs; long-range systems; optical systems

---

## 1. Introduction

The tasks of comparing and image search are the main tasks in computer vision. The quality of the search, low sensitivity to distortions - the fundamental requirements for search algorithms. There are many approaches to solve such tasks, as well as a wide range of technical solutions to the problems of range-finding using lasers and other systems. At the same time there are a number of problems that imply the impossibility of active far-range systems use. In the described project is planned to develop and implement effective methods of image sections searching based on characteristics of pixel's local neighborhoods.

The relevance of the project is ensured by the need to develop methods for solving labor-consuming probabilistic-geometric problems of digital image processing. The complexity of these tasks is growing in connection to information and computing technologies development. New research methods of solution and specific applied problems are combined to the modular principle of components of all the systems being created. It makes possibilities of successful implementation of investigating problematic and applied problems.

## 2. Current state and features of the range-finding systems

The task of determining the distance to the object is extremely urgent in such areas as geodesy, military science, navigation and computer vision. Rangefinders are used to determine the distance.

Range-finding devices are divided into active and passive. The initial development of the range-finding devices was among the passive systems but in recent years the most widespread got active range-finders due to the simplicity of their implementation, their unpretentious use and the rather high accuracy of measurements.

The principle of active type rangefinder functioning consists in time measuring spent by the sent signal from the rangefinder to the object and back. The speed of the signal propagation is considered as known. Also there are active range-finders estimating changes in the parameters of the reflected signal (phase or power).

It should be noted that there are a number of problems in which the use of active range finders is difficult or not possible. These tasks include using of rangefinders on low-visibility platforms. The use of emitters leads their unmasking as well as long-distance measuring require installation of high-power emitters that could be dangerous for users.

Measurement of distances by passive range finders is based on determining the height  $h$  of an isosceles triangle  $ABC$  on the known side  $AB = l$  (stereo-base) and opposite acute angle  $b$  (so-called parallactic angle). At small angles  $\beta$  (expressed in radians)  $h = l/b$ . One of the quantities,  $l$  or  $b$ , is usually a constant, and the other is a variable (measured). By this feature, rangefinders with a constant angle and range finders with a constant base are distinguished. In general, based on the distance between the observation points  $l$  (stereo-base) and the angle of displacement  $\alpha$ , the distance to the object is calculated:

$$h = \frac{l}{2 \sin \frac{\beta}{2}}$$

For a long time, the use of stereovision principles in systems has not been considered in practice due to the relative complexity of implementation and poor quality of digital images. In recent years, the quality parameters of shooting equipment increased significantly, which suggests the possibility of developing such systems. In addition, the stereovision system can possess in addition to low-visibility (due to the implementation of the passive calculation technique) also by a functional allowing the measuring on post-production. It is possible to use stereo-pairs and recalculate the parallax for a multitude of image points in the overlapping area and the sharply displayed image space in comparison with active range finder systems which allows to obtain the distance-range information only at the time of shooting.

Based on the above facts, it seems extremely relevant to develop both theoretical methods in the field of stereo-range metering, as well as technical hardware-software solutions.

### 3. Passive systems of range-finding

A digital image obtained with a passive stereo system carries only color or brightness information and does not possess any additional data.

Images of the real world include a narrow set of colors or luminances, and therefore, when solving the problem of determining conjugate identification is not for individual points, but for fragments of images. Thus, it is extremely important for a point taken in one image to know, where at the second image is its conjugate and how to compare these fragments correctly.

A comparison of the neighborhoods of conjugate points does not yield to strict formalization. It is based on the problem of identifying images of fragments of the three-dimensional world from images. This task can hardly be adequately described in a formal way. Significant differences in views lead to the appearance of projective and brightness distortions when shooting. It is of fundamental importance that these differences depend not only on the geometry of shooting-system, but also on the geometric and physical characteristics of the surface itself. The location of the light source influences to the surface affects the light distribution. The position of the surface elements and their properties determine the amount of energy that enters the camera lenses and the local differences in the brightness of the conjugate fragments of images.

The significance of the differences depends on the difference in the viewing angles. The more this difference (in particular, the larger base), the less similar the images becomes. Therefore, all methods of comparing neighborhoods of conjugate points rely more or less on a formal approach rather than on the character of images. Also important is the possibility of their preliminary processing, reduction to an epipolar stereopair, construction of efficient descriptors of neighborhoods of conjugate points for accuracy and speed for comparison.

In an idealized situation, the values of the similarity function in the scanning process along the line should represent a one-moment peak value for the desired pixel when the zero similarity value is returned for all other pixels (neighborhoods) of the line.

During processing real graphics data, such combination of values returned by the function of similarity is impossible. But processing initial data with sufficient information to identify a local region, the graph of the function retains a sufficiently clear extremum, which allows to identify the desired pixel of the image.

The main task in the construction of the disparity map is the selection of a variant for comparison of regions in which the extremum of values of the similarity function will be most pronounced. This involves defining for the point some characteristics that would uniquely characterize the point of the image. Moreover, the conjugate point on the reference image had identical or maximally similar values of similar characteristics.

The final step of calculating disparity is the aggregation of total or averaged values. When aggregating, as the resulting disparity for the point  $p$  of the base image, the value of  $d$ , is chosen, where the minimum value of the cost is reached. In the real situation, it is possible to find several values of  $d$  with the same or minimum values of disparity (especially when averaging the values). The problem of possible ambiguity arises in most methods of constructing disparity maps, which is associated with optical, mechanical, electronic features of cameras. A solution to multi-valuedness can be the introduction of certain conditions on the value of disparity, for example, the largest, average or smallest possible.

### 4. Combined method of stereo reconstruction

Usually, in practice, measures are taken based on the sum of the absolute differences or the sum of the squares of the differences. Both functions (summation over a given window) allow you to calculate the cost effectively enough when the corresponding pixel of the conjugate image has the closest intensity value. These functions are extremely sensitive to the quality and parameters of the original data (exposure, glare, overexposure, underexposure, matrix noise, random emissions).

In the process of carrying out a computational experiment on real-world images, it was determined that correlation methods of comparing local parts of images give stable results on textured areas and extremely low accuracy in homogeneous areas (there are no explicit contrasts). During analyze of "alive" systems, we can conclude that in nature the distance to a homogeneous object is also poorly localized and its binding to the boundaries due to reflex saccadic movements.

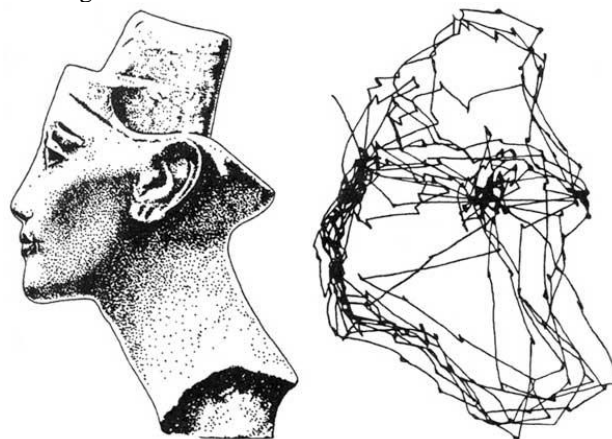


Fig. 1. Recording the movement of the eye (scanning while viewing the head of Nefertiti) according to Yarbus, 1965.

It can be argued that a comprehensive approach to the problem of ranging is required, including both consideration of the possibilities of increasing the uniqueness of the means of identification, and the development of a combined approach using different techniques to the neighborhoods of points, depending on their local characteristics.

Work with digital stereo images allowed us to determine the following combined method: correlation areas with the maximum uniqueness of points should be applied to areas of images that have contrast objects in their area (brightness differences), and to homogeneous ones - to bind to remote contrast boundaries by building a set of vectors on several directions.

Thus, in the preprocessing phase, it becomes important to compile a calculation map based on the proximity to the contrasting areas. In order to classify a point as being on a brightness difference, the brightness change associated with a given point must be substantially greater than the change in brightness at the background point. In connection with the specifics of local calculations, the way to determine the "essential" values to establish a threshold. In turn, the concepts of the first and second derivatives are used for the quantitative expression of the brightness variation.

The definition of an image point as a drop point occurs if its two-dimensional derivative of the first order exceeds a certain predetermined threshold. The calculation of the first derivative of a digital image is based on various discrete approximations of a two-dimensional gradient. The direction of the gradient vector coincides with the direction of the maximum rate of change of the function  $f$  at the point  $(x, y)$ .

$$\begin{array}{ccc} z1 & z2 & z3 \\ z4 & z5 & z6 \\ z7 & z8 & z9 \end{array}$$

The calculation of the gradient of the image consists in obtaining the values of the partial derivatives  $Gx = df/dx$   $Gy = df/dy$  for each point. One of the methods for finding the first partial derivatives  $Gx$   $Gy$  at a particular point is to apply the following gradient Sobel operator:

$$\begin{aligned} Gx &= (z7 + 2 * z8 + z9) - (z1 + 2 * z2 + z3) \\ Gy &= (z3 + 2 * z6 + z9) - (z1 + 2 * z4 + z7) \end{aligned}$$

It is necessary to determine the appropriate masks for the Sobel operator, which identifies horizontal and vertical contours (brightness differences) for convolution with the original image. It is also possible to change the above formulas that give the maximum response for contours directed diagonally. Additional pairs of Sobel's masks for detecting gaps in diagonal directions can be defined as:

$$\begin{array}{ccc} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{array}$$

for points lying on the diagonal edge -45 degrees;

$$\begin{array}{ccc} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{array}$$

for points lying on the diagonal edge +45 degrees.

For each of the masks the sum of the coefficients equals zero. That means that these operators will validate zero response on the areas of constant brightness, which is characteristic of the differential operator. The masks considered are used to obtain the gradient components  $Gx$  and  $Gy$ . To calculate the magnitude of the gradient these components must be used together:

$$Gradient = |Gx| + |Gy|$$

Calculation map is to be constructed on the computed gradient map base. This process implies the recognition of the area as low-textured in the event that in the search window around the point less than 15% of the area is occupied by contrasts.

Direct calculation of disparity occurs in several stages. At the first stage, high-textured sections are processed using the correlation functions of similarity measures, such as, for example, Euclidean distance, cross-correlation, etc.

In the world practice only brightness information is usually used as criteria for comparing image points. The disadvantage of this approach is the color interpretation multiplicity for points with the same brightness value. In addition, one should take into account the fact that the perception of color and monochrome images is uneven. This feature is taken into account in the methods of degradation of the color model of the image to 256 shades of gray due to the introduction of coefficients applied to the respective channels.

$$Y = 0.299 * R + 0.587 * G + 0.114 * B$$

Most of the images were initially formed by a color sensor in color. Therefore, in order to increase efficiency, the use of color information is seen as obvious.

Working with three components of color can be represented as a "cloud" of points in three-dimensional space with the axes corresponding to the color channels of the image. However, the RGB space is not orthogonal, due to the specificity of the human visual analyzer, which has a different number of rods and cones that are susceptible to a particular color.

Since the correlation function of a three-dimensional space is a measure of correlation functions that use the Euclidean distance, which is correctly calculated in orthogonal systems, one should perform orthogonalization of the space RGB into the space XYZ.

The representation of the RGB base colors, according to the ITU recommendations, in the XYZ space has the following correction factors:

$$\begin{aligned} Red: x &= 0,64 \quad y = 0,33 \\ Green: x &= 0,29 \quad y = 0,60 \\ Blue: x &= 0,15 \quad y = 0,06 \end{aligned}$$

Therefore the transformation system for translating colors between RGB and XYZ systems can be represented in the following form:

$$X = 0,431 * R + 0,342 * G + 0,178 * B$$

$$Y = 0,222 * R + 0,707 * G + 0,071 * B$$

$$Z = 0,020 * R + 0,130 * G + 0,939 * B$$

After reduction of spaces, operations that are valid for orthogonal systems to points can be applied. Using the "color" image processing increases the potential uniqueness of the point 1.72 times (the maximum distance between the luminance values in the gray scale is 255 units, color values 416 units).

After this step is performed the distances from the points in the low-textured regions to the nearest contours in several directions are calculated. A group of multidimensional characteristic vectors is formed, to which a similar approach is applied, as well as to vectors with color information.

As the stages are completed, the disparity map is filled. Due to the fact that all operations are in strict accordance with the calculation card, auto-aggregation of the results of different stages into a single map occurs.

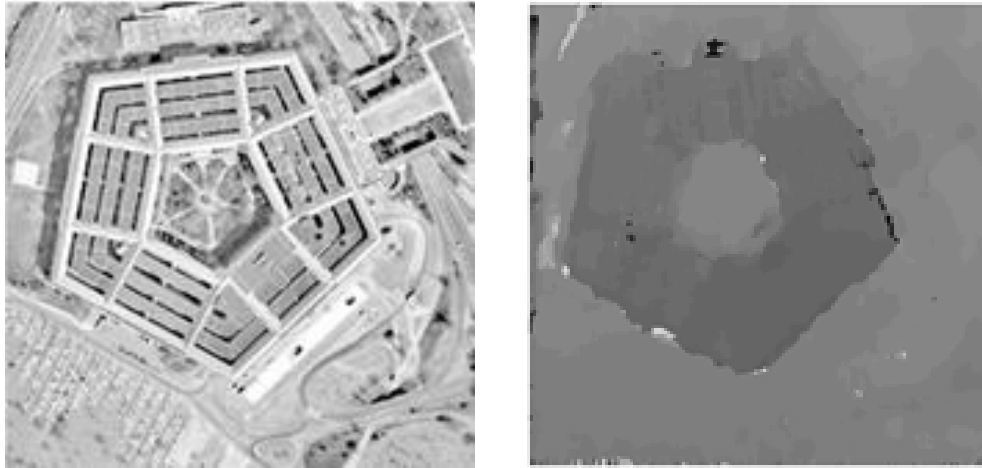


Fig. 2. Image of stereopair and map of disparity.

Despite the multi-stage implementation, this algorithm has a large number of cyclic stereotyped locally independent operations, which makes it possible to parallelize the algorithm.

## 5. Calculation of the distance to the object on the basis of heterogeneous initial data

Generalized the principle of determining the position of points in space on the basis of disparity data has been repeatedly described in the literature. Suppose two cameras  $L$  and  $R$  are installed in such a way that their  $X$ -axes are collinear, and the  $Y$  and  $Z$ , axes are parallel. The centers of the cameras are displaced relative to each other by an amount  $b$ , corresponding to the base of the stereoscopic system. When observing a certain point of the space  $P$  the point  $P_l$  is formed on the left image, and on the right  $P_r$ .

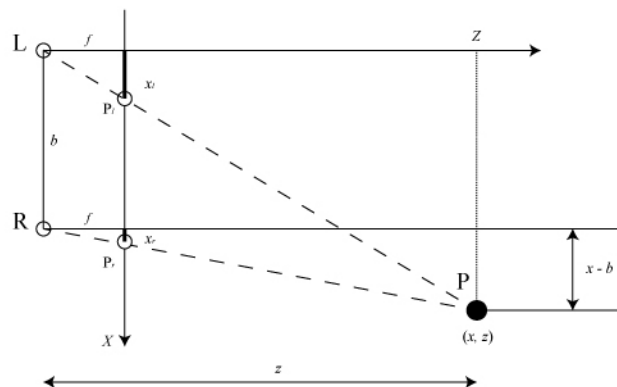


Fig. 3. Geometric model of a stereoscopic system.

Considering the similarity of two pairs of triangles, we obtain the equations:

$$\frac{z}{f} = \frac{x}{x_l} \quad \frac{z}{f} = \frac{x - b}{x_r} \quad \frac{z}{f} = \frac{y}{y_l} = \frac{y}{y_r}$$

It should be noted that by construction, the coordinates of the image points  $y_l$  and  $y_r$  can be considered the same, which corresponds to the rectified system with a rigid connection between the photosystems (3D camera, human visual system). Given this property, it is possible to transform the system of equations for the explicit expression of the coordinates  $x$ ,  $y$ ,  $z$  of  $P$  in real space on the basis of the coordinates of the projections of points on stereopair images:

$$z = fb/(x_l - x_r) \quad x = \frac{x_l z}{f} = b + \frac{x_r z}{f} \quad y = \frac{y_l z}{f} = \frac{y_r z}{f}$$

The solution of the system of equations allows one to uniquely calculate the position of a point in space.

Unfortunately, the given system of equations is not applicable for digital reconstruction, because there is a mixture of different systems of dimension: disparity value in pixel distance, focal length and base in metric units. However, it is possible to calculate the distance to the point using the disparity value, the angle of the lens alignment and the base of the stereo system.

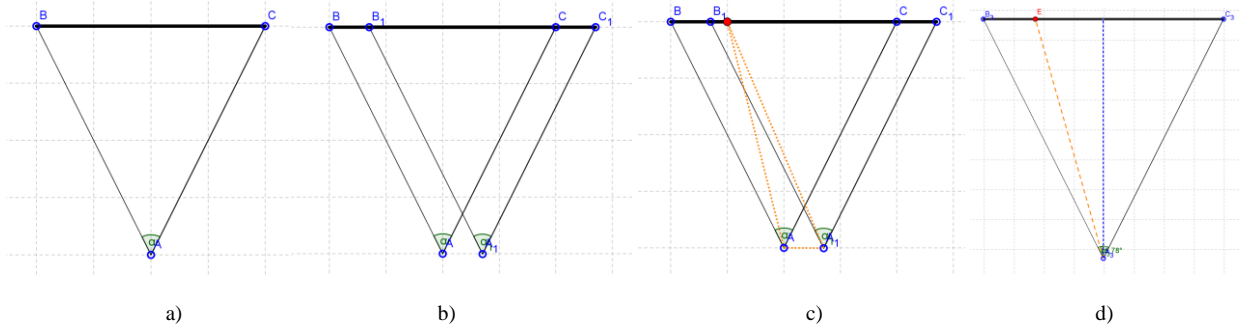


Fig.4. Geometrical model of calculating distance on the lens alignment.

The stereovision system can be represented in the following form:

- A, A<sub>1</sub> – observation point;
- BC – left image;
- B<sub>1</sub>C<sub>1</sub> – right image;
- BC<sub>1</sub> – zone of overlap;
- α – horizontal lens opening angle.

The calculation of the distances to the object is the problem of solving the triangle B<sub>1</sub>AA<sub>1</sub>.

Within the system, the base of the triangle is known - the base of the stereo system. Angles at the base can be calculated through the angle of the lens opening.

The calculation of the viewing angle to the object of interest for the left camera is performed as follows (Fig. 4c):

$$\angle EAA_1 = \beta = 90 \pm \arctg\left(\frac{\operatorname{tg} \frac{\alpha}{2} * \left(\frac{BC}{2} - BE\right)}{\frac{BC}{2}}\right)$$

Where:

- β – angle of the desired triangle;
- α – angle of the horizontal lens opening;
- BE – the X- coordinate of the image point;
- BC – the width of the image (Fig 4d).

Знак «+» используется в системе при  $BE < \frac{BC}{2}$ , «-» при  $BE > \frac{BC}{2}$  соответственно, а при  $BE = \frac{BC}{2}$  принимаем  $\angle EAA_1 = \beta = 90$

The angle of sight is calculated in the same way as the correction for the fact that the sign «-» is used in the system for  $BE < \frac{BC}{2}$ , «+» for  $BE > \frac{BC}{2}$  and for  $BE = \frac{BC}{2}$  and  $\angle EA_1A = \beta_1 = 90$ .

The third angle can be obtained by the formula:

$$\beta_2 = 180 - \beta - \beta_1$$

By the sine theorem, it is possible to determine the lengths of the sides A<sub>1</sub>E and AE

$$AE = AA_1 \frac{\sin \beta}{\sin \beta_2} \text{ and } A_1E = AA_1 \frac{\sin \beta_1}{\sin \beta_2}$$

Due to the possible inclination of AE relative to the horizontal plane of the system, it is necessary to bring EA into the plane of the system

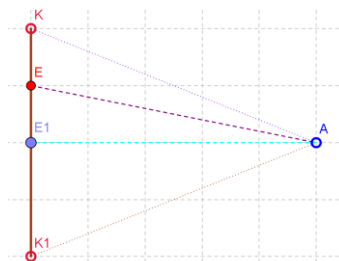


Fig. 5. Geometrical model of vertical declination of the system.

In situation of  $KE < KK_1/2$

$$\vartheta = \arctg\left(\frac{\operatorname{tg} \frac{\alpha}{2} * \left(\frac{KK_1}{2} - AE\right)}{\frac{KK_1}{2}}\right)$$

In situation of  $KE > KK_1/2$

$$\vartheta = \arctg\left(\frac{\operatorname{tg} \frac{\alpha}{2} * \left(AE - \frac{KK_1}{2}\right)}{\frac{KK_1}{2}}\right)$$

In situation of  $BE = BC/2$

$$\begin{aligned}\vartheta &= 0 \\ \vartheta &= \angle EAE_1 \\ AE_1 &= AE \cos \vartheta\end{aligned}$$

As a result, the distance from the reference (left) camera to the object and the angle relative to the base (plane of the matrices) of the system are obtained.

It should be noted that increasing the measured distance increases the sensitivity of the system to the accuracy of the alignment of the system and the quality of the images. Since the angles at the base of the system take values close to  $90^\circ$ . This leads to the fact that the values of trigonometric functions change very dynamically.

## 6. Conclusion

During the research, the author made a study of the existing methods for processing stereo images in the tasks of stereo reconstruction. The algorithms functioning patterns are revealed and the causes of their unstable work are determined. A combined image processing technique that takes into account the characteristics of local image sections is proposed. Additionally author examined the problem of the heterogeneity of the initial data necessary for obtaining metric information of three-dimensional objects in the field of interests.

The algorithm developed and described was implemented by the author in the form of a program library, which can later be used in a wide range of applications. Due to the fact that the matrix of distances to image points can be translated into a specific coordinate system for one or another application system. It should also be noted that the organization of the user's access to the functions allows for more flexible use of the library.

In addition, the described technique has found its application in a number of software and hardware and software developments that are carried out at the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. Specifically, as element of the mobile topogeodetic system and as program library for ERS system.

## References

- [1] Borodach A, Tuzikov A. Automatic determination of matching points on two images. Proceedings of the 9th International Conference "Pattern Recognition and Information Processing", 22-24 May, Minsk, Belarus 2007; 1: 49–53.
- [2] Shapiro L. Computer vision. Moscow: BINOM. Laboratory of Knowledge, 2006; 752 p. (in Russian)
- [3] Volkovich AN. Use of color characteristics in the construction of disparity maps. Materials of the International Congress of ROPI-2011. Nizhny Novgorod: UNN 2011; 64: 112–117. (in Russian)
- [4] Lyakhovsky VV, Volkovich AN, Zhuk DV, Tuzikov AV. Sistem automatic reconstruction of three-dimensional scenes for several images. Materials of the V Belorussian Space Congress, October 25-27, Minsk 2011; 2: 129–133.
- [5] Zhuk DV, Tuzikov AV. Reconstruction of a three-dimensional model using two digital images. Informatics 2006; 1: 16–26.
- [6] Shulgovsky VV. Fundamentals of Neurophysiology. URL: <http://www.braintools.ru/rubric/information/from-books/fundamentals-of-neurophysiology> (01.02.2017).