Construction of a 3D depth map from binocular stereo.

Conference Paper · February 1994

1 author:

Wojciech Mokrzycki
Cardinal Stefan Wyszynski University in Warsaw

Wojciech Mokrzycki
Institute of Computer Science, PAS,
01-237 Warsaw, Ordona 21,
e-mail: wmokrzyc@plearn.bitnet,
fax: (++48-22) 37-65-64

CONSTRUCTION OF A 3D DEPTH MAP FROM BINOCULAR STEREO

The goal of 3D machine vision system is to analyze visual data acquired from the environement (scene) and carry out its geometrical measurement or derive an interpretation, to execute given task. Construction a 3D depth map of the scene is a first and main subtask in each task executed by the 3D machine vision system.

This paper presents some results (obtained in recent years), concerning the construction of the depth map, based on an image stereopair registered by a binocular camera system.

Key words: binocular stereo, feature extraction, stereo matching, 3D depth map.

1 Introduction

Typical task of 3D machine vision system concerns the navigation of autonomus vehicles, the inspection of manufactured parts, the analysis of microscopic images, etc. The goal of the system in these applications is to measue and describe or identify and locate a specified object in the scene (i.e. to determine its exact location and orientations). If the system should identify or recognize an object, it must have full knowledge about its shape. Such a priori knowledge about the object is provided with the aid of a model of this object, which contains the geometrical data of the object and, may be, some other data, e.g. photometric, thermal. A vision system which uses objects models is refered as a **model-based, recognition vision system** (fig. 1). In the other case the system, which ónly"describes the scene in terms of image primitives, is referred to as a **mapping vision system**.

A high quality recognition system should be able to locate objects when:
- objects have arbitrary shapes and forms,
- objects are viewed from any direction,
- objects are partially ocluded by other objects.

To design such system one must solve many problems:
(1) the method of data acqusition,
(2) the method of construction of the object models,
(3) the means of description the acquired data and the model,
(4) the method of matching the data descriptions and the objects descriptions. Their general

classification it can find in [21]. In this paper I consider only vision systems based on the stereo image pair acquired from binocular stereo system of cameras.

There are two main approaches to the construction of **objects models** (for model-based - recognition system):
In **the first** approache the actual objects are used to generate a model, i.e. data obtained from many viewpoints are collected into a coherent form to provide information about the object from all the viewing points.
In **the second** approache a CAD system is used with the set of graphic primitives which allows the user to construct interactively the models of the objects (most recent research efforts use CAD-like systems). The **representation scheme** for the acquired data and the model object - is a key issue in the 3D computer vision, since the 2D arrays of numbers from the cameras are not convenient to use in their żaw"form. Must be used to describe the data and the models. In order to describe the data and the models, the representation scheme has to be:
- unambiguous, (i.e. no two objects have the same representation),
- unique, i.e. there is a single description exist for each object using the representation scheme,
- robust, i.e. with respect to missing data points, e.g. as in the case of partial occlusions,
- convenient to use in the matching and storing.

The **matching** of the appropriate descriptions of the data as well as the models is usually performed in two steps:

1. A correspondence is estabilished between the two sets of descriptions usually between the partial description of the object and the full model description; the correct match **estabilishes an interpretation** of an input data,
2. On the base of the estabilished correspondences, geometrical transformations, usually in the form of rotation matrix and the translation vector are derived to give the possibility of the transformation of the data base model into the orientation of the object in the scene.

2 Acquisition of visual data

From many existing methods of data collection about a scene, I will consider acquisition of the stereo pair of images with the aid of passive, binocular stereo system of cameras, and stereo analysis. Two cameras form the **stereo system** if they are such oriented, that their view areas have common part. The stereo analysis relates to that common view area.

The **foundation of stereo analysis** presumes that the scene objects contain certain structures, transformed into stereo image features like edges, common areas, unique shapes. In the stereo analysis these features are extracted (from both stereo images), and the images are mutually matched and it a śtereo disparity"is calculated. The **stereo disparity** is the difference $(<x_l, y_l> - <x_r, y_r>)$ between the coordinate projections of a scene point P(X,Y,Z) on the image planes: O_l (left) and O_r (right) of the stereo pair.

Stereo analysis consists of the following stages, [4]:

1. **stereo pair acquisition**; stereo images can be collected in many ways; however if the analysis must be exact, the differences in the images of the stereo pair should result only from different positions of the cameras in the camera system (and not from, e.g. differences in parameters or differences in an acquisition time in the both cameras),
2. **camera modeling**; dependences among X,Y,Z (the coordinates of some scene point), kno-

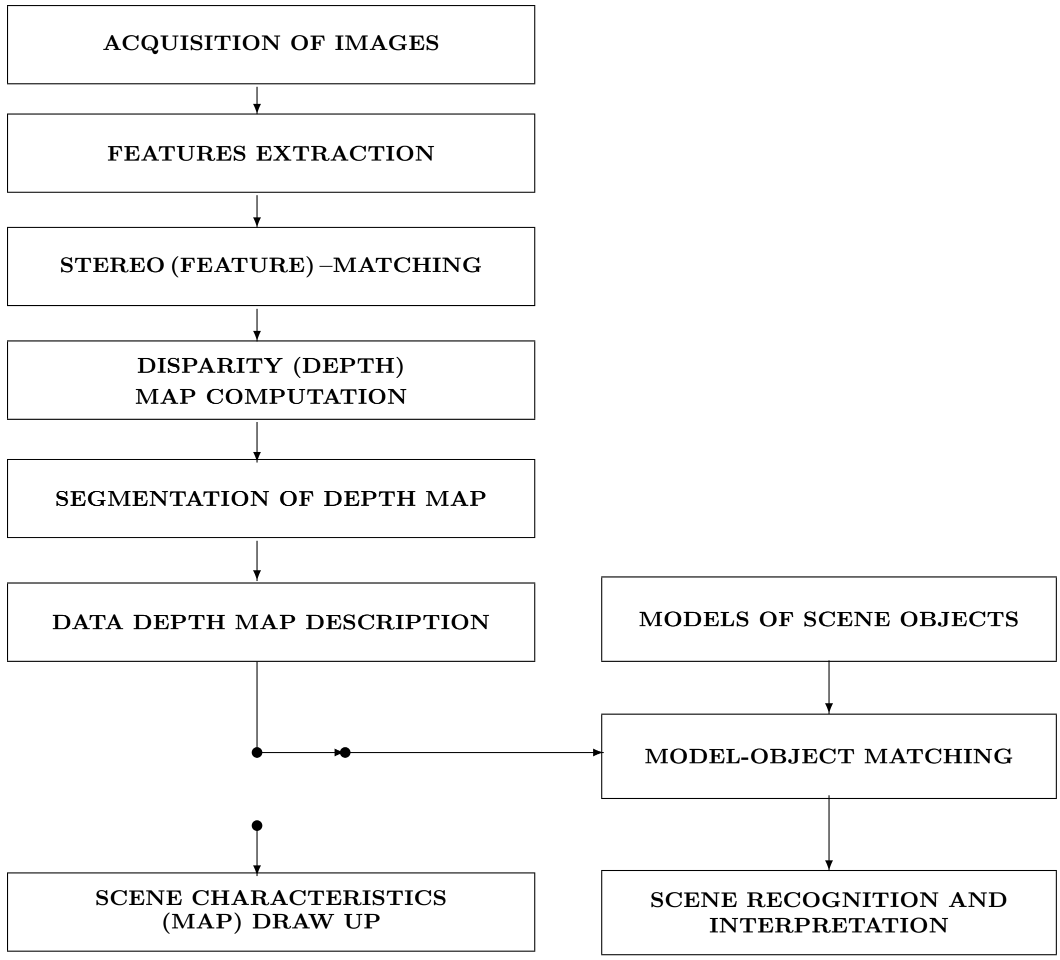

Fig.1. A general scheme of mapping and recognition in the model-based stereo vision system.

wing coordinates x,y (the projection of those point on the image plane), geometry and optics of the camera system are estabilished,

3. **feature extraction and image segmentation**; feature-based and area-based processing is carried out to find **distinctive matching primitives**,
4. **stereo matching** - finding of the distinctive points and structures in the scene; for both views of the scene; a correspondence must be estabilished between those point that are visible in both images,
5. **disparity and depth maps construction**; the distance between the matched points coordinates on both images (disparity map), and between matched points on the images and the suitable points in the scene (depth map) are to be calculated,
6. **interpolation**; it is often necessary to interpolate values for those regions, for which no disparity can be found.

2.1 Base dependences in lateral stereo camera system

In the stereo camera system there is so called **epipolar geometry**. This geometry includes, among others, [8], the terms: ępipolar planeąnd ępipolar lines”.

The **epipolar plane** contains certain point P(X,Y,Z) of the scene and its projections $< x_l, y_l >$ and $< x_r, y_r >$ on the camera image planes O_l and O_r. The epipolar plane cuts the camera image planes along lines called the **epipolar lines**.

From above it follows the ȩpipolar geometry constraint":
evry point on the epipolar line on the left/right image has the corresponding point on the appropriate epipolar line on the right/left image, fig.2. The above epipolar geometry constraint plays , in stereo analysis, very important role. In the stage of stereo matching this dependence allows to reduce the searching for the corresponding points on the coupled image from entire image plane to the appropriate epipolar line; this is important reduction of stereo matching complexity.

Although the positions and the orientations of the cameras (in the stereo camera system) can be arbitrary, nevertheless (for operating possibility of the stereo system), the common (for the both cameras) view area should be the largest. It means that the y_l and y_r axes of the stereo image coordinate systems should be parallel and the z_l and z_r axes should be converged (at least, to be parallel). If the cameras are oriented in such a way that the y_l and y_r axes are parallel, we say that **cameras are in correspondence**. If the cameras are in correspondence, the epipolar lines are parallel to picture scanning lines and only horizontal disparity can be on the stereo images (i.e., $y_l = y_r = y$ and $< x_l, y_l > - < x_r, y_r > = < x_l - x_r, 0 >$).

2.1.1 Stereo cameras system with parallel axes

When the axes of the cameras systems coordinates are parallel, fig. 3, the following formulas are true:

$$x_l = \frac{f(X + b/2)}{Z}, \quad x_r = \frac{f(X - b/2)}{Z}, \quad y_l = y_r = \frac{fY}{Z}. \tag{2.1}$$

Defining the stereo disparity as:

$$d \triangleq x_l - x_r, \tag{2.2}$$

the backprojection equations have the form:

$$X = \frac{b(x_l + x_r)}{2d}, \qquad Y = \frac{by_l}{d}, \qquad Z = \frac{bf}{d}. \tag{2.3}$$

Constraints along epipolar lines

We can use the above formulas to calculate the scene point coordinates (if we know the image coordinates x_r and y ($= y_l = y_r$) of the corresponding points on the stereo images, obtained by the stereo matching process).

In the stereo camera system with the parallel axes, the view area of the both cameras does not cover each other, fig. 4. The x_r coordinate of the P_r point (right image), reaches a minimum, when $P(X,Y,Z)$ scene point reaches the SB_r line (the right border of the common view area). In the same situation the x_l coordinate of the P_L point (left image) reaches minimum. These minimal values we can calculate with the aid of the following formulas:

$$x_{rmin} = \frac{b - h}{2} + fb/l_c, \qquad x_{lmin} = -\frac{b + h}{2}, \tag{2.4a}$$

where l_c is an assumpted maximum of the depth of the scene exploration.

The x_r and x_l coordinates of the P_l and P_r points, adequately, reach the maximum, when the $P(X,Y,Z)$ scene point reaches the SB_l line - the left border of the common view area. These maximal values we can calculate with the aid of the following formulas:

$$x_{rmax} = \frac{b + h}{2}, \qquad x_{lmax} = -\frac{b - h}{2} - bf/l_c. \tag{2.4b}$$

Also, the inequality below is evident:

$$x_r > b + x_l. \tag{2.5}$$

The above formulas estabilish additional constraints on the images coordinates of the corresponding points along every epipolar lines. The constraint on the Z coordinate value (the minimum of the scene depth), is:

$$Z_{min} = f + l_s, \tag{2.6a}$$

where f is the focus lenght of the camera and l_s can be calculated from the formulas below:

$$l_s = (b/2)\text{ctg}\sigma. \tag{2.6b}$$

2.1.2 Stereo camera system with converged axes

The strategy and correctness of the consecutive stages of the stereo analysis, particulary the matching process, and especially operating possibility of the stereo system need (in many tasks) the view areas of both cameras cover entirely each other. It can reach it if the cameras axes become converged, fig. 5. How large should the angle of the camera convergence be? It depends on answers on some other questions:.
(1) how large should the maximum depth of the scene exploration be?
(2) How large can the precision of the data acquisition be?

Let us assume that the resolution coefficient w[mm/r.u.], is the line size of the (square) area of the scene, represented by the one (square) pixel of the image. Then we can calculate an a - the line size of the (square) view area of the camera in the following way:

$$a = rw, \tag{2.7}$$

where r – the line size of the image (in r.u. - a raster unit). We assume, the areas views of the both cameras cover each other. Now, having estabilished (for many other reasons), the lenght of the stereo base b, we can calculate the angle of the convergence of the cameras, using the below formulae:

$$\theta = (1/2)\arcsin\frac{4bfh}{rw(4f^2 + h^2)}. \tag{2.8}$$

For the stereo camera system with the axes converged on 2θ, we can calculate the X, Y and Z coordinates of the $P(Z, Y, Z)$ scene point, in the followinf way:

$$= \frac{b(f\cos\theta - x_l\sin\theta)(f\cos\theta + x_r\sin\theta)}{fd\cos 2\theta + (f^2 + x_l x_r)\sin 2\theta}. \qquad X = \frac{bf(x_l + x_r)}{2(f^2 + x_l x_r)\sin 2\theta} \qquad Y = \frac{by_l}{d}. \tag{2.9}$$

The l_c value follows from the converge angle value (and the technological parameters of the cameras), and can be calculated as follows:

$$l_c = \frac{b(\cos 2\theta + \cos 2\theta)}{2\sin 2\theta} \tag{2.10}$$

The second constraint on the calculating Z coordinate value (the scene depth - the vertex of the view cone), $Z_{min} = f + l_s$, can be calculated in the following way:

$$l_s = (b/2)\text{ctg}(\sigma + \theta) = (b/2)\frac{2f - h\text{tg}\theta}{h + 2f\text{tg}\theta}. \tag{2.11}$$

The parameters dealing with the image acquisition are influencing the quality of the results of feature extraction (regarding to the image contrast and the signal-noise-ratio) as well as the quantity of the extracted features (regarding to the number of sensor elements of the camera and regarding to the sampling frequency used for the analogue image signal).

3 Image segmentation and stereo matching

It is not suitable to evaluate a selected single processing step in stereo vision, e.g. a stereo matching algorithm, without taking into consideration the interdependence of all processing steps. The whole stereo vision cycle has to be examined and has to be assessed. However, at the heart of a binocular stereo approach lies the task of stereo matching. Stereo matching has a great influence on the quality of the computed results. In addition, the entire correlation and the mutual dependence of the processing steps within a stereo system should be taken into consideration.

3.1 Feature- and area-based segmentation

The most commonly used features are **points along the edges of intensity discontinuities**. These points, which are termed **edgels** for edge elements, are useful because they represent the points at which high-confidence, anambiguous matches may most likely be made. However, the feature-based methods provides only sparse matches and require interpolation as well as some methods for modeling occlusion. Large local change in disparity in the feature-based process may confuse it.

Area-based methods have been applied, where the surfaces varies smoothly and continuously. They offer the adventage of directly generated the dense disparity map but are sensitive to noise and breakdown when there is a lack of texture or when depth discontinuities occur.

Some preconditions for the feature extraction method are already fixed by selecting a special technique for stereo matching (matching of: pixels, line segments, regions, etc.).

3.2 Constrains and rules for stereo matchig

The base for solving the correspondence problem is **similarity**. In general for a particular image feature in one image, there will be many candidate matches in the corresponding image. Matching rules derived from the constraints underlying the physical environment and imagining are used to restrict this pool of candidate matches.

The common constraints incorporated in stereo processing, are the following, [6, 16]:

1. **surface continuity constraint**: it assumes that the physical world is composed of surfaces being almost continuous everywhere; this suggests that disparities ought to vary smoothly;
2. **surface uniqueness constraint**: this relates to the fact that the imaged surfaces, for the most part, are opaque, and there is assumed that the image element acquired by a camera corresponds to a unique point lying physically on the surface of an object. Thus correspon dence should be unique;
3. **general position constraint**: this relates to the observation that certain events occur quite infrequently, in a statistical sense, to rule out false correspondences.

They where identified by two categories of stereo matching rules, [6, 16]:

I. spatial-domains rules: the rules are based on the surface continuity and the general position assumptions of the matching environment; examples are the following:

*** area statistics**: matching primitives collected over an image measurement patch are compared across images to obtain a single similarity measure; as an example can be correlation of image feature; this rule usually implies a strong continuity assumption, because it imposes

approximately constant disparity over the patch;
* **contour statistics**: in this the comparison is restricted along a contour; the assumpion of physical surface continuity is made weaker by assumimg disparity to be smooth along contours but allowing them to change abruptly across contours; figural continuity is an example of this type.

II. gradient limit rules: the rules are based on the manner in which images are manifested; examples are the following:
* **ordering constraints**: impose the restriction that along epipolar lines the matched primitives must occur in the same order; this is equivalent to the assumpion that imaged surfaces are not transparent and are continuous;
* **disparity gradient limits**: restrict the maximum disparity gradient allowed between mached primitives; there has suggested that for most natural scene surfaces the disparity gradient between correct maches is usually less than one;
* **coarse-to-fine analysis**: disparity information obtained at a coarser scale is used to limit the search domain for the matching of finer scale primitives; this is used with scale specific matching primitives.

3.3 Feature-based stereo matching

The feature-based approaches match more abstract features rather than matching texture regions in the two images, since such features are less sensitive to noise. Feature-based analysis provides more precise positioning (for the feature) in the individual images, and it can attain correspondingly higher accuracy for its correspondences in 3D.

Below there are following feature-based matching methods used the most frequently:

1. A method using a local disparity limit, [17]. Line segments, which are defined by the zero crossings of the Laplacian of the smoothed image, are matched with regard to their differences in orientation when their centers of gravity lie inside a specific search area. The search area is defined by using local disparity limits computed by applying the block matching technique.
Constraints: epipolar geometry, uniqueness and local disparity limit.

2. Marr, Poggio and Grimson multi-resolution approach, [2, 5]. The algorithm is based on investigations about the human visual system. Candidates are matched with regard to the minimum difference between the directions of the zero- crossings in the Laplacians of the smoothed images in different scales.
Constraints: epipolar geometry, uniqueness, compatibility and disparity continuity.

3. A statistical method as suggested by Kass, [10] . A vector of 12 "uncorrelated" functionals is defined by the first and second partial derivatives of the smoothed intensity function with three different standard deviations. Pixels are matched if the differences between all vector components are smaller than the respective user defined thresholds.
Constraints: epipolar geometry and uniqueness.

4. A method using disparity continuity as suggested by Kim and Bovik, [13]. **Extreme points** detected in a contour (that are high curvature edge points terminations and junctions) are matched in a first processing step. Afterwards the remaining contour points are matched with regard to the smallest value of the disparity gradient. The contours are defined by the zero crossings in the direction of the local gradient of the smoothed intensity function.

Constraints: epipolar geometry, uniqueness, disparity continuity and disparity gradient.

5. A method using disparity histograms, suggested by Shirai and Nishimoto, [9]. Candidates are matched with regard to global and local disparity histograms. The distribution of possible disparity values in a local neighbourhood of a matching candidate is computed as disparity histogram in different scales and different window sizes. The best scale is selected with regard to the most significant peak in the disparity histogram. The disparity with the greatest likelihood is computed and compared regarding its compatibility with neighboured candidates.
Constraints: epipolar geometry, uniqueness, compatibility and disparity continuity.

6. A method using dynamic programming as suggested by Ohta and Kanade, [8]. Connected edges are matched by finding the shortest paths in 2-D and 3-D search areas. The problem of finding a matching path on a 2-D search plane, whose axes are the right and left scanlines, is called intra-scanline search. The search in a 3-D search space, which is a stack of the 2-D search planes, is used to utilize the consistency constraints across the 2-D search planes. This search is called inter-scanline search.
Constraints: epipolar geometry, uniqueness and ordering constraint.

3.4 Area-based stereo matching

The area-based matching is more accurate than those using edge-based primitives since regions have higher discriminations capability. Also, area-based approaches are more efficient since there are fewer features to be matched. However, while the use of regions make stereo matching more accurate, reliable, and efficient than edge-based matching, region-based matching processes typically yield coarse disparity maps. The most frequantly mentioned matching methods are:

1. Shirai stereo matching method, [12]. The local minima of a similarity function based on the square differences between the intensity values in search windows of different sizes, are determined. Pixels are matched starting from the smallest window and using three thre sholds.
Constraints: epipolar geometry and uniqueness.

2. Block matching for stereo, [17]. The similarity between the intensity distributions in two equal sized blocks ($n \times m$ matrices) in the left and the right image is measured by using the mean square error MSE between the intensity values of the pixels inside the respective blocks. The left image is segmented into a constant number of equal sized blocks and the search for a corresponding block in the right image is only carried out for these segmented blocks. The same disparity value is assumed for all pixels of one block when applying the block matching technique.
Constraints: epipolar geometry and uniqueness.

4 Depth map interpolation and depth discontinuities location

Deterministic approaches

In both feature- and area-based correspondence methods, the interpolation is usually included into the normal processing. In addition to interpolating, there are used the smoothing surfaces and isolating depth discontiuities. Terzopoulos, [14], attempt to locate discontinuities by locating significant **inflection points** on the resultant surface. Blake and Zisserman, [11], introduced their **graduated nonconvexity** algorithms, which allow the direct search for depth discontinuities and orientation discontinuities, respectively.

Saint-Marc et al., [18], present a method to smooth the surface while preserving discontinuities, which facilitates the detection of discontinuities.

Any one of these interpolation methods may be used to extract the discontinuities in the manner that we suggest by adding a strong preference for the existance of the discontinuity contour (either depth or orientation) to occur at edgels or along the edges.

3D (control) points for the interpolation of the visible surfaces can be computed exclusively for the matched features in both images.

Hoff and Ahuja, [15], attempt to combine the feature matching, contour detection, and surface interpolation into one process. Their results are very impresive, but they fail when their matching features (zero crossings) are too sparse; in addition, they can not accurately locate the surface discontinuities.

Stochastic approaches

Geman and Geman, [7], use Gibbs distributations to model the spatial correlations between neighboring regions within images along with the use of an additional cost constraint (the łine process") that represents the presence or absence of a surface discontinuity.

Stochastic approaches are, however, computationally expensive in a typical implementation and contain nonrobust parameters that need specyfic tuning to each scene. Deterministic approaches are more adventage and robust with respect to their internal parameters.

5 Combined stereo vision processing

Marr and Poggio, [2], proposed a computational model of human stereo vision, which may be the base to machine vision system bulding.

In this model zero crossings in the Laplacian of the Gaussian of the intensity, are used in stereo matching. They suggest, that three constraints should be satisfied in choosing global correspondence: compatibility, uniqueness, and continuity. Grimson, [3], implemented an improved version of this model and obtained good results when there is a sufficiently dense set of features. However, zero crrosings, usually, are relatively sparse and irreguraly distributed in images, so there is produced a sparse depth map. Such methods must be augmented by an interpolation step.

The generation of accurate and finer resolution disparity maps (and subsequently depth measurements) can be better accomplished using edge-based techniques. Both regions and edges play important, but somewhat complementary role in a binocular stereo process. The new stereo vision system is investigated which, unlike most stereo approaches, integrates ąrea-basedąnd "feature based"primitives (fig. 6, see also [19]). These area-based processing provides a dense disparity map, and the feature-based processing provides an accurate location of discontinuities. In this way it is possible to generate a disparity map that is sufficiently accurate to allow to detect depth and surface orientation discontinuities, provided that the resolution is fine enough. Area-based cross correlation along with an ordering constraint and a weak surface smoothness assumption allow to produce an initial disparity map. However, a match is accepted only if both views agree on a correlation peak and if this peak is strong enough. This disparity map is only a blurred version of the true one because of the smoothing introduced by the cross correlation.

At depth discontinuities the problem can be reduced by introducing the edge information: the disparity map is smoothed (subject to the constraint that the disparity at edgels is fixed) and the unsupported points removed. This metod gives an active role to edgels parallel to the epipolar lines.

Description of the method

First, the original images are adjusted so that their corresponding epipolar line lie along corresponding rasters.
Next, the resulting images are reduced by a convolution with a Gaussian and subsampled by a factor of two to form a pyramid of three image pairs, each of which is separately processed starting with the coarsest pair.

The initial feature- and area-based processes proceed independently to produce a set of edge features and a dense disparity estimate. These are then combined to produce a dense disparity estimate with less blurring.

At each level of the image pyramid this estimate is used to improve the matching in the next finer level, and the last set of results are used the surface features: specially labeling all of the points as being visible or not, providing a confidence for the generated values, and marking the **depth discontinuity contours**. Further processing may then be performed to smooth the discrete surface , interpolate through unknown areas, and extract the **orientation discontiuuities**. This system operates in several passes over the data.
First, a set of well-scattered, reliable matches are obtained by **locating interest points** based on variance and edge strength and then utilizing an unconstrained hierarhical match algorithm.
Next, a camera calibration is performed , and epipolar-constrained hierarhical matching algorithm is used to **match the interest points**. Those points that are evaluated to have the most reliable matches are used as **anchors** for a final matching of all of the interest points. The results and experience reached that way can be integrated into a solution with regard to the selected requirements. The methods will be selected concerning their methodical distinction in solving the correspondence problem (area based and feature based, etc.). These stereo methods have been evaluated with regard to their suitability for the generation of a precise geometry description of 3-D objects.

6 Conclusion

Many research activities dealing with stereo vision are known. Nevertheless, there still does not exist a standardized way for the evaluation of the algorithms. Usually many questions occur concerning the evaluation of published solutions to a problem. The known methods differ in their solution to the segmentation method or correspondence problem as well as in the selection of constraints assumed for the visible objects. Furthermore, nearly all publications usually present their own solution in the task domain rather different, without comparing it to the results of other methods (e.g. measurement, robot vision, photogrammetry, stereo microscopy, etc.). The large number of distinguishable features in the solutions aggravates a direct comparison of the methods and it is nearly impossible to evaluate the suitability of a method for a selected application.

The aim of this research is to take advantage of the ideas and the results of the published methods. An exact reimplementation is not possible anyway for different reasons, e.g. computer architectures, programming languages, details etc.

The realization of this framework is motivated by the idea of applying some selected algorithms to the combining stereo vision process, giving the possibility to build autonomously acting mapping stereo vision system. The combination of a framework with a testbed for image acquisition allows for a systematic comparison between different methods. Therefore, the suitability of a method regarding to a selected application can be assessed directly.

The considerations mentioned above encourage the necessity to create an experimental tool for the methodical investigation of computer vision methods. Another difficulty occurring with the evaluation of stereo methods is resulting from the direct interdependence between the single processing steps. Furthermore, the processing steps are influenced by the selection of the constraints on the objects in the scene. The accuracy of the results representing the geometry of an a priori unknown 3-D object is the one of the main criterions in the evaluation.

References

1973

[1] Mori K., Kidole M., Asada H.:An iterative prediction and correction method for automatic stereo comparision. CGIP, 2, 393–401

1979

[2] Marr H.P., Poggio T.: A computational theory of human stereo vision. Proc. Roy. Soc. Lond., B204, 301–328

1981

[3] Grimson W.E.I.: From Images to Surfaces: A Computational Study of the Human Early Visual System.Cambridge, MA:MIT Press

1982

[4] Barnard S.T., Fischler M.A.: Computational stereo. ACM Computing Survays, 14(4), 553–572

1983

[5] Grimson E.: An implementation of the computational theory of visual surface interpolation. CVGIP, 22, 39–69

[6] Nishihara H.K., Poggio T.: Stereo vision for robotics. First Int. Symp. Robotic Res., IEEE Press, 489–505

1984

[7] German S., German D.: Stochastic relaxation, Gobbs distribution, and the Bayesian restoration of images. IEEE trans. PAMI, 6(6),721–741

1985

[8] Ohta Y., Kanade T.: Stereo by intra and inter-scanline search using dynaming programming. IEEE Trans. PAMI, 7(2), 139–154

[9] Shirai Y., Nishimoto T.: A stereo method using disparity histograms of multi-resolution channels. 3rd Int. Symp. on Robotics Research, Gouivieux, France, 27–32

1986

[10] Kass M.: Computing visual correspondence. In: From pixels to predicates: recent advances in computational and robotic vision. Pentland A. (Ed.), Ablex Publ., Norwood, New Jersey, USA, 78–92

1987

[11] Blake A., Zisserman A.: Visual Reconstruction. Cambridge, MA: MIT Press

[12] Shirai Y.: Three-dimensional computer vision. Springer, Berlin

1988

[13] Kim N.H., Bovik A.: A contour-based stereo matching algorithm using disparity continuity. PR, 21(5), 505–514

[14] Terzopoulos D.: The computation of visible-surface computation. IEEE Trans. PAMI, 10(4),417–438

1989

[15] Hoff W., Ahuja N.: Surface from stereo: integrating feature matching, disparity estimation, and contour detection. IEEE Trans. PAMI, 11(2), 121–136

[16] Marapane S.B, Trivedi M.M.: Region-based stereo analysis for robotic applications. IEEE Trans. SMC, 19(6), 1447–1464

1991

[17] Koschan A.: Stereo matching using a new local disparity limit. In: Computer Analysis of Images and Patterns. Proc. 4th Int. CAIP'91 Conference, Dresden, Germany, 1–26

[18] Saint-Marc P., Chen J.S., Medioni G.: Adaptive smoothing: a general tool for early vision. IEEE trans. PAMI, 13(6), 514–529

1992

[19] Cochran S.D.,Medioni G.: 3-D surface description from binocular stereo. IEEE Trans. PAMI, 14(10), 981-994

[20] Mokrzycki W.: Stereo vision systems. A review of problems. Machine GRAPHICS & VISION, 1(1/2), 342-392

1993

[21] Mokrzycki W.: Stereometry of 3D scene. Proc. II Symposium TPO'93, Serock, 18-20 Nov. 1993 (in Polish)

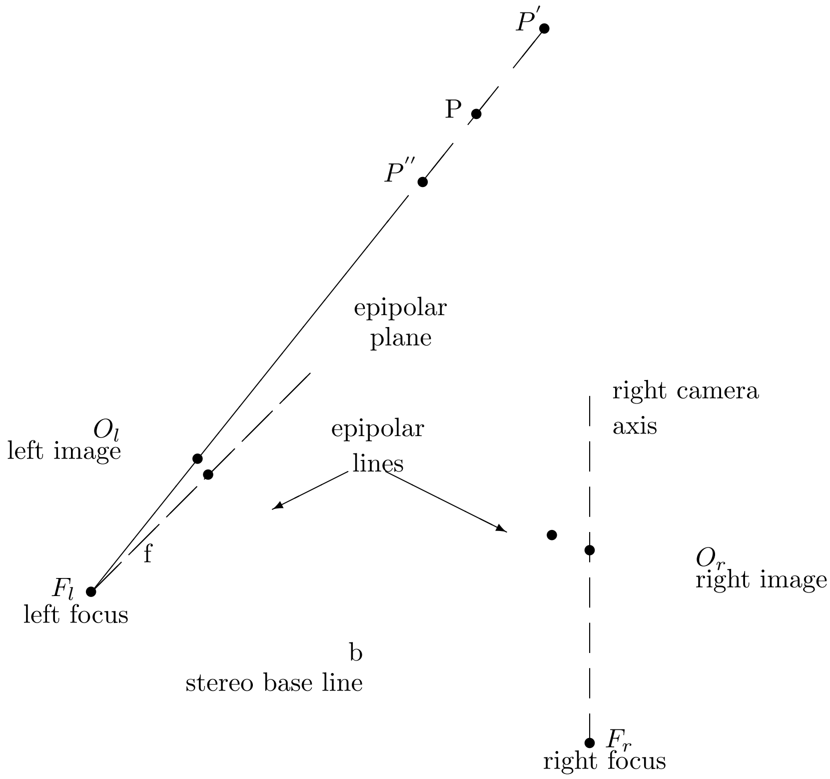

Fig.2. Epipolar geometry of stereo camera system

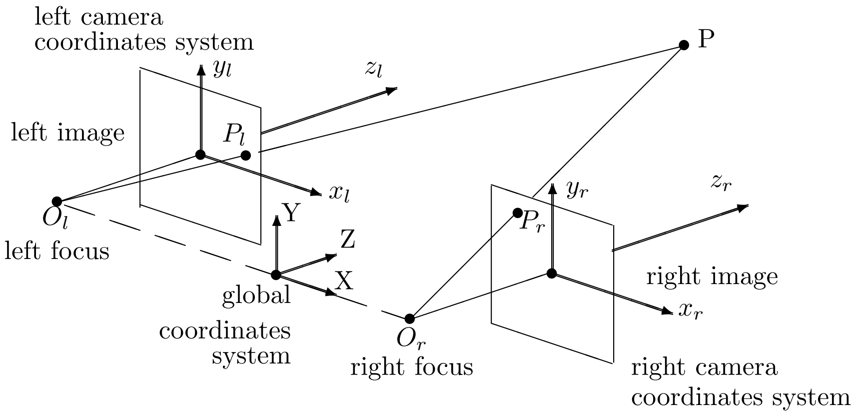

Fig.3. Coordinates systems: global (of the scene) and of the cameras images

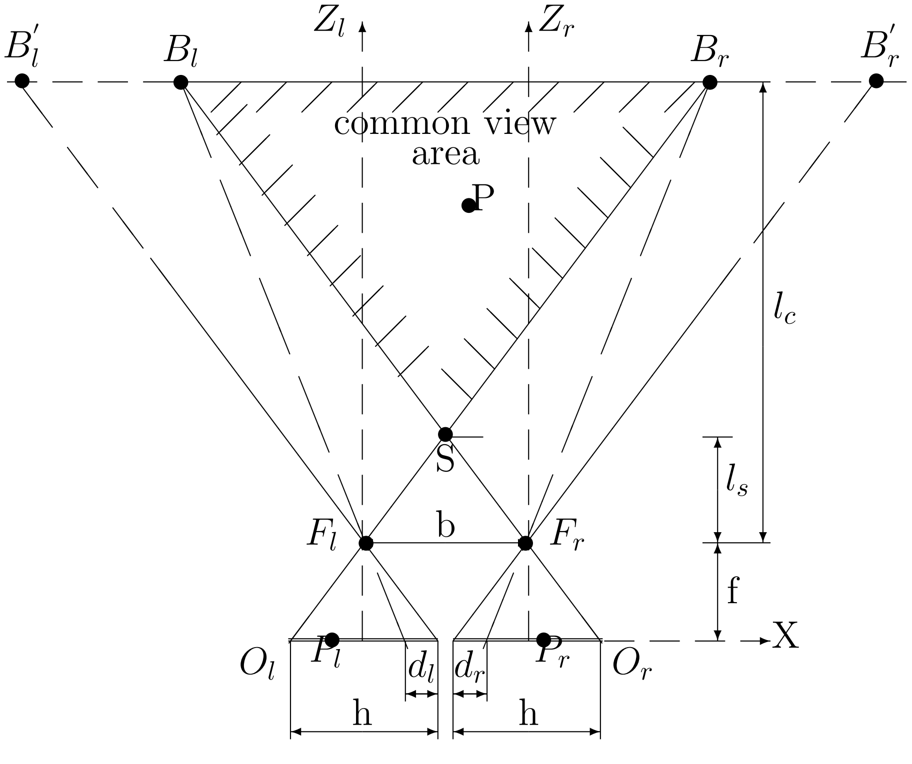

Fig.4. Areas of cameras views

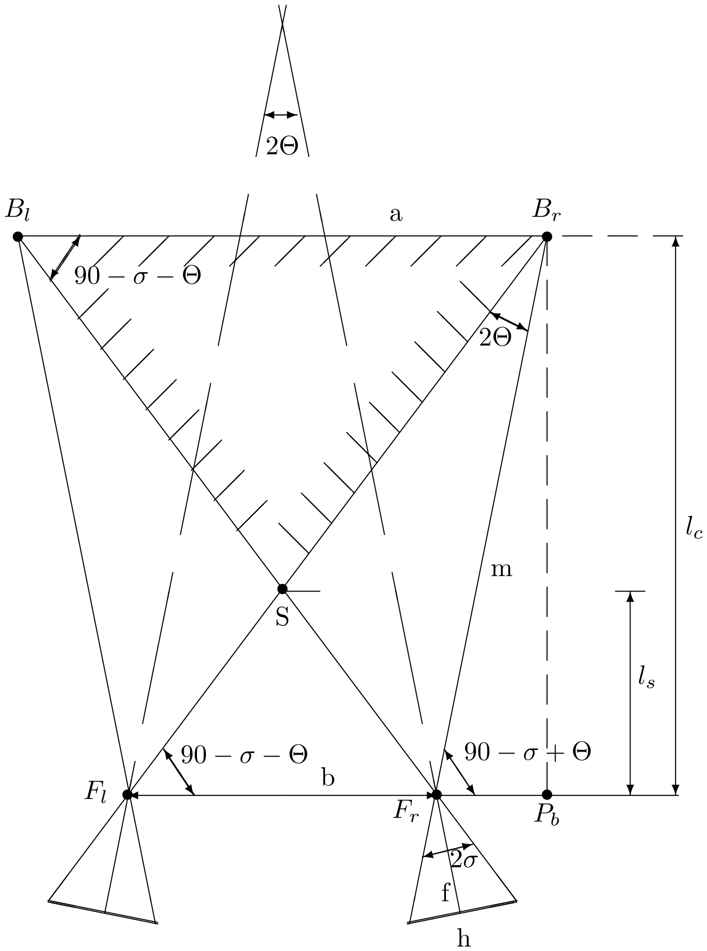

Fig. 5. Dependences in convergence cameras system

Fig.6. A combined stereo processing flow diagram