Some techniques of real-time disparity estimation

Vladimir N. Kolmogorov, Ivan A. Matveev, Alexander B. Murynin

Computing Center of RAS

Moscow, Russia

Abstract

One of the crucial problems in machine vision is reconstruction of 3D shape of objects from their images and the most complicated aspect of this problem is disparity estimation. Some techniques of disparity calculation in real-time monocular and binocular environment are presented. The developed algorithms are based on the correlation method of calculating corresponding points on images in stereo set. Performance optimization is achieved through some modifications of the correlation method, including pyramidal representation of the images. The method for estimating size and shape of the scanned region is proposed. The system was tested on a database of stereo-images.

Keywords: *computer vision, stereoscopic vision, correlation algorithm, pyramidal data representation*

1. INTRODUCTION

Rapid evolution of computer vision systems makes it actual to develop effective methods of surface characteristics recovery from images for real-time applications. The basis for this recovery is disparity estimation between images of an object observed by a computer vision system. Here a method for disparity estimation based on the correlation algorithm is presented. This method is developed for application in person recognition system [2], as well as the modification of principal component analysis described in [3]. The first section presents some background overview of the problem and certain restrictions to the methods used. In the second section the general approach to the solution of the surface reconstruction problem using the correlation method is described. The next section presents the optimization techniques that allow taking advantage of specificity of the problem. The final section gives the example of practical application of the method and the results obtained.

Various stereo surface reconstruction algorithms for different applications are reviewed in [5][6]. Applicability of any of these methods to a particular problem depends substantially on the equipment used, the scene geometry, computing resources and many other factors. Thus, of the great number of published works in this field only few are applicable to the problem of reconstructing complicated surface from low-contrast images. Stereo-reconstruction methods can be classified firstly by the number of source images used (i.e. points in space-time from which the registration is performed). Thus, these methods can be divided into monocular [7], binocular [8][9][10] and multi-ocular [11]. Secondly, they can be classified by the approach to surface reconstruction as those defining surface local slopes using brightness modulation, those using texture and combined ones. The algorithm presented is binocular and performs texture-based reconstruction.

The problem of texture-based stereo-reconstruction in binocular systems was formulated and treated from physiological point of view by Marr who modeled human vision. Basing on this Grimson [13] built a computing system and showed the efficiency of this model. Marr and Poggio developed the model still further by introducing the principle of pyramidal data representation and treatment [4]. Pascal Fua applied stereo-reconstruction methods to images of human face [12].

2. BASIC CONCEPTS

For reconstructing 3-D surface a series of successive images of a moving object is used. The series of images is divided into a set of pairs. Each pair of images is treated as a virtual stereopair and is processed using the correlation algorithm to calculate disparity maps [1]. The basic concepts of the correlation algorithm and its peculiarities related with object motion are discussed here.

The problem is to reconstruct the 3D-shape of some object given two images of this object obtained from different points of view. Here some definitions and notions of presented framework are introduced. Let us call left image *L-image*, right image *R-image* and their aggregate *stereo-pair*. The points on L- and R- images representing the same point of a body surface are called *corresponding* points. Having imposed some frame of reference on these images, one can say that *disparity at a point of left image* is the measure of the distance between this point and the corresponding point of right image. Generally, one may write:

$$L(x, y) = R(x - D_X(x, y), y - D_Y(x, y)) + N(x, y)$$

where D_X and D_Y are projections of disparity at point (x, y) of right image to OX and OY axes, L(x, y) and R(x, y) are left and right images and *N(x, y)* is a function that includes noise and changes in image that originate from variations of light and shading conditions. It is likely, that an analytic solution of this equation cannot be found. Thus, to calculate disparity one can use several numerical methods.

It is easy to see that given the geometry of the system (i.e. orientation and distance between cameras) one can use coordinates of correspondent points to restore the position of original point in 3D-space.

In order to obtain good relief estimation, with no gaps or inaccurately reconstructed areas, one should use a dense map of correspondent points (also referred to as *disparity map*). So, the question is how to find correspondent points for as many points of the image as possible. The first thing one can think about is to treat sequentially all points of, say, left image and find their correspondents on right image. This is a correct but highly time-consuming procedure. So a number of optimized techniques are developed, which are presented in the next section. Let us consider the question how to find a correspondent point for a given one. Let us call the image, for point *(x, y)* of which a correspondent is searched, *the basis* image, and the image, on which the correspondent point is searched *the scanned* image. Some region in the scanned image, where correspondent point can probably be found is called *the scanned region*. Some small neighborhood of the point is called *the correlating region*. The correlating region is built (on the basis image) and then the most resembling one is sought in the scanned region of the scanned image. Measure of distance between regions' centers on the basis image and those on the scanned image is considered to be a disparity in this point.

$$D(x,y)=\rho((x,y),(x',y'))$$

Now the question is how can one measure the resemblance of regions i.e. what is a correlation function. The well-known form of the correlation function is

$$\Re_1(f,g)=\sum_{x_i\in\Omega_1,y_i\in\Omega_2} f(x_i)*g(y_i)$$

But it is applicable to some stochastic models, while the nature of our problem is different. It is difficult to derive theoretically the suitable form of the correlation function. So several other forms of correlation functions were tested, for example:

One can see that the canonical correlation function increases with the increase of similarity while others presented behave in an opposite way. But still they measure the similarity, if properly normalized. Now the correspondent point can be estimated as:

where $\tilde{\Re}$ is the normalized correlation function. Below $\Re$ will be referred to as *correlation function* for short since $\tilde{\Re}$ can be easily computed from $\Re$.

Since the algorithm is based on comparison between regions of images it is reasonable to make some signal pre-processing to facilitate this comparison and make it more reliable. The algorithm is texture-based so it's easy to conclude that the best pre-processing would emphasize texture details. Methods employed for extracting local features of images including texture are based on local window operations. For this purpose an approach based on local equalization is chosen. It is known that equalization results in strengthening such space variations of signal that have typical scale size of nearly same magnitude as the equalization window. In other words the equalization process selects in the best way details that have nearly the same size as the equalization window. Thus, selecting the size of equalization window one can achieve a distinct emphasizing of cardinal details of the image. Large-scale variations emergent due to irradiance trend along the image frame are suppressed.

3. OPTIMISATION TECHNIQUES.

The algorithm presented is primarily developed for the case of motion. Since the movement of the object is quite arbitrary this case is much more difficult than the case of stereo. The computational complexity in this case is determined by several factors.

1. In order to process N images we should run correlation algorithm for approximately N stereo pairs.

2. Disparity can lie in both X and Y direction; so instead of a rectangle stretched in X direction as in the case of stereo the scanned region should be a square of the same linear size.

3. The initial values of disparities in X and Y directions are unknown.

4. The dispersion of disparity on the image can be small or large depending on the kind of movement of the object. For example, in the case of parallel transfer the dispersion is quite small, whereas in the case of rotation around OZ axis the dispersion is much higher. Thus, the problem of choosing the best size of the scanned region, optimal from the points of view of calculation time and required accuracy, arises.

In order to overcome these difficulties several optimization techniques are proposed.

These techniques substantially use the assumption that the surface of an object is sufficiently smooth and thus it becomes possible to use different smoothing filters and interpolations without considerable loss of precision.

The image of an object can have regions with well-marked texture as well as those with feebly marked texture and even textureless regions. That is why from the point of precision it would be optimal to choose combined algorithm, which uses texture and correlation approach in the regions with well-marked texture and shading in the regions with feebly-marked texture. But the variants of such algorithms available today are rather time-consuming and do not satisfy time constraints for real-time recognition systems. Oscillations of relief are little if compared to the distance from an object to the camera

Now we shall enumerate the problems that led to algorithm restructuring. One of the substantial limitations is the run-time of an algorithm. In order to decrease the run-time the *pyramidal data representation and workflow* are used. In some points due to noise or absence of texture or too regular texture the algorithm can fail and a wrong disparity value can be output. To sift out these errors the *reverse pass* of correlation algorithm and median filters are used. One of the critical issues is the choice of the scanned region. The method for its automatic choosing, based on the statistical analysis of data received from calculations performed with small number of points and small scanned region, is proposed. Then a method allowing to substantially decrease the calculation time of the correlation function is presented. Finally the method for finding maximum of the correlation function is discussed.

3.1 Pyramidal representation

As it was already mentioned the system involved was designed to work in real-time mode. Let us consider some optimization steps that can be performed in the framework of the algorithm discussed. If the number of points in which the correlation search of correspondents is performed is denoted as N, number of points in the scanned region as P and the square of the correlation region as L an estimation will give us the following calculation complexity:

$$n \sim N * P * L \quad (1)$$

where n is the number of elementary operations. So the problem of diminishing the calculation complexity can be reduced to a problem of decreasing any of these numbers without affecting precision of surface reconstruction. There is one simple and effective way of doing it. The mesh of points, where the correlation search of correspondents is performed is thinned N times, so the number of these points is N^2 times less. (One should notice that such calculation is not equivalent to calculation upon an image lessened N times for such image transformation removes information about thin image texture). Then the values of disparity in other points are interpolated. Linear interpolation by nearest neighbors is feasible at moderate thinning (N<4). At greater thinning interpolation is made through constructing of smooth surface, containing known points of the thinned mesh. However this approach is not used since at great thinning the precision is lost at points that are not in the mesh while at low thinning the calculation procedure is rather time-consuming. One of the ways of removing this contradiction is pyramidal structure of data and processing. The trick is to split the processing into two or more similar steps (or layers). Characteristics relating to large fragments of images are calculated earlier on higher levels of pyramid and each lower level involves characteristics of smaller image fragments using the information obtained at higher levels. The system described employs two-level pyramid. Rough disparity that is calculated at high level is mainly used to lessen the scanned region at low level correlation search. Also the size of the correlation region is decreased at low level. Let us consider each of the two levels in details.

High level. The mesh of points where the correlation search of correspondents is performed is substantially thinned. Since the results of this calculation are used in the following steps, a wrong value at one point of mesh can affect many points of resulting disparity. That is why both calculation and check (i.e. reverse pass described below) are performed thoroughly, which means that large correlation regions and large scanning areas are used. If a value in a point fails the reliability check the disparity is calculated in one of nearest points of image. This value of course can slightly differ from that in the required point but such small differences are treated at lower levels of the algorithm. If none of nearest points gives reliable values of disparity all points in the neighborhood up to other points of the mesh are marked as unreliable. The lower level will not use the information of high level at these points (i.e. the lower level will work as thoroughly as the higher level does). The substantial complexity of calculations at each point at high level does not significantly influence the total calculation cost since the number of points in the thinned mesh is small. After disparity in all points of the mesh is calculated, disparity in all other points is interpolated by a smooth surface. Here the assumption is used that the surface of the object is smooth enough so the interpolation surface is a good approximation of it by Chebyshov's norm (i.e. there are no points in which real and interpolated disparity strongly differ).

Low level. Disparity is calculated on a slightly thinned mesh with exception of the points that were treated at high level. Since some approximate value of disparity is already present at any point as a result of high level efforts the region where the corresponding point is located (the scanned region) can be determined more accurately, so it can be made smaller. Since the scanned region is diminished, the number of regions locally resembling the correlation region decreases as well, so no false correspondent may appear. Therefore the correlation region can be lessened too. In some cases (if high correlation is obtained) the check via reverse pass can be omitted. So then, with the help of the information obtained from high level all three multipliers in (1) are decreased.

3.2 Checking disparities.

Let us now concentrate on possible mistakes of the correlation algorithm. If a studied surface has a highly regular structure several regions can be found in the scanned image that resemble a certain region of the basis image. In this case correlation function has several local maxima in the scanned region, and it may happen, that thy ‘correct’ maximum is not a global one. So, a wrong match can be selected for a point and wrong disparity can appear. The other case, when correlation algorithm can generate an error is the absence of good texture. In this case the correlation function has no maximum, any point in the scanned region can be correspondent and disparity can receive any value. To suppress errors in the case of texture absence a simple but still rather effective technique is used. Some threshold is imposed on correlation function in such a way that if it’s value does not exceed the threshold the regions are considered to be completely different and are excluded from further treatment. It is much more difficult to eliminate errors in the case of highly regular structure. To reveal such errors the reverse pass of the correlation algorithm is introduced. Suppose for a given point (X_L,Y_L) in L-image the correlation algorithm has found a possible corresponding point (X_R,Y_R) in R-image. After this to ensure the perfect correspondence the reverse pass technique is applied, which is: just the same correlation algorithm treating R-image as the basis and L-image as the scanned image. If this procedure finds that the correspondent for (X_R,Y_R) is the initial point (X_L,Y_L), the correspondence is considered perfect. Otherwise the point (X_L,Y_L) is marked as having unrecognized disparity which is taken into account during following steps of algorithm.

The assumption about the smoothness of the surface can also be employed for eliminating incorrectly found points. The transformation $D \to D'$ mapping the points of the left image to the points of the right image and the inverse transformation $D' \to D$ are smooth enough. This makes possible using the following simple algorithm for checking found disparities. The disparity (dx_0, dy_0) of the point (x_0, y_0) is considered to have been found incorrectly and eliminated from further consideration if the disparity does not satisfy one of the two median filters – for direct and for inverse transformations:

1.

where ($\langle dx \rangle$, $\langle dy \rangle$) is the average disparity of the points adjacent to (x_0, y_0) except (x_0, y_0) or

2.

where ($\langle dx^{rev} \rangle$, $\langle dy^{rev} \rangle$) is the average disparity of points which was projected to the point adjacent to $(x_0', y_0') = (x_0, y_0) + (dx_0, dy_0)$ except (x_0, y_0).

(the points (x_1, y_1) and (x_2, y_2) are adjacent if $|x_1-x_2| \le \delta$, $|y_1-y_2| \le \delta$ where δ is the mesh granularity).

This simple technique allows to correct errors of the correlation algorithm at all levels of the pyramidal algorithm practically without computational expenses.

3.3 Determining the size of the scanned region.

As indicated above the dispersion of disparity on the image can change greatly from one stereo pair to another. That is why the size of the scanned that can provide certain accuracy can range in some diapason. Thus the problem of determining this diapason arises.

The proposed method is the following. First some small initial scanned region P_0 is set. Then the correlation algorithm is run for some number of points large enough to carry out statistical analysis. After that by analyzing the histogram $h(P_i, dx, dy)$ it is determined whether the scanned region was set correctly or it should be enlarged. In the latter case a new scanned region P_{i+1} is calculated and the previous step is repeated. To make the algorithm computationally efficient all calculated values of the correlation function may be preserved and used in successive steps.

Experiments show that the histogram for the large scanned region covering the correct values of disparities is usually a well-clustered two-dimensional area P_{max}. The figure [1] illustrates a typical 2D histogram for approximately 600 points.

To find the area P_{max} the following algorithm consisting of two steps is proposed.

Step 1. On this stage there is no information at all about the location of P_{max}. The algorithm starts with the small scanned region of square form with the center at (0, 0) and successively increases the size of the scanned square. If the scanned region does not cover P_{max} found disparities are randomly located in the scanned region; when the scanned region starts covering P_{max} they begin to cluster. Thus we need the reliable criterion to determine whether the histogram $h(P, dx, dy)$ contains the clustered set. Several criteria were tested and the following one provided the best results.

	12	13	14	15	16	17	18	19	20	21
-3			1							
-2									1	
-1				1	1		1			
0			4	3	8	4		1		1
1			3	20	21	16			1	1
2		4	9	32	38	26	13	4		
3	1	4	19	55	52	40	26			
4		4	17	30	50	15	12	5	2	
5			2	4	7	13	8	4		
6			1	1		2	1			
7						1				
8										

Figure 1 2D disparity histogram

We will say that points (dx', dy') and (dx'', dy'') from P are *δ-connected* if there exists such sequence of points from P $(dx_1, dy_1) = (dx', dy')$, (dx_2, dy_2), …, $(dx_n, dy_n) = (dx'', dy'')$ that $h(P, dx_i, dy_i) > 0$, $1 \le i \le n$, and $\rho((dx_i, dy_i), (dx_{i+1}, dy_{i+1})) \le \delta$. The relation of δ-connectivity is reflexive, symmetric and transitive. Hence there exists a partition of P according to this relation. Let P_0' be the set from this partition containing the maximum number of points. Let us denote *δ-closure* of the set P' from P as

$$[P'] = \{(dx, dy) \in P \mid \\ \exists (dx', dy') \in P' : \rho((dx, dy), (dx', dy')) \le \delta\}$$

It is obvious that δ-closure of a δ-connected set is a connected set. Now we test the hypothesis that $P_0 = [P_0']$ is the sought clustered region. For this purpose we introduce two hypotheses - H_0: $p(P_0) = p_0$ and H_1: $p(P_0) > p_0$ where $p(P_0)$ is the probability for disparity to get into P_0 and p_0 is this probability in the case of the uniform distribution: $p_0 = |P_0|/|P|$. Let n be the number of points in P and n_0 is a variate corresponding to the number of points in P_0. H_1 is accepted if $n_0 \ge n_{0min}$ and is rejected otherwise. We want to minimize the probability of type I error – the probability that H_1 will be accepted if H_0 is true. Hence n_{omin} may be determined from the condition that the probability of type I error is quite small (say less than or equal 0.05):

$$\alpha = \sum_{k=n_{0\min}}^{n} C_n^k p_0^{\,k} (1-p_0)^{n-k} \le 0.05$$

Thus the criterion for determining whether there is a clustered set can be formulated as follows. For $\delta = 1, \dots, \delta_{max}$ find maximum δ-closure $P_{0\delta}$ and calculate $\alpha_\delta = \sum_{k=n_0}^{n} C_n^k p_0^{\,k} (1-p_0)^{n-k}$; if for some δ $\alpha_\delta \le 0.05$ then we decide that the clustered set is found.

The typical number of points involved in this step is about 50.

Step 2 (high level of the pyramidal algorithm). The aim of this step is to calculate disparities on the thinned mesh; these values will be used in the low level of the pyramidal algorithm. From the previous step we have only an estimation of P_{max} which can be used as the initial approximation. Since P_{max} is a well clustered connected set it is reasonable to suggest that neighbors of points from P_{max} for which many points were found also lie in P_{max}. Thus the following algorithm may be used for determining P_{max}. We start with the scanned region found in the previous step. The procedure for choosing the new scanned region P_{i+1} is as follows:

the point (dx_0, dy_0) will be added to P_{i+1} if the number of points in its Δ-neighborhood exceeds some fraction from the total number of points n:

When $P_{i+1} = P_i$ we proceed to the low level of the pyramidal algorithm.

3.4 Algorithm for calculation of the correlation function.

Let us suppose that the correlation function is to be calculated for a large number of points N of an image in the scanned region P of the same shape and size for all points. The following correlation function will be used:

$$\Re_3(x_0, y_0, dx, dy) = \sum_{(x,y)\in\Omega} | f(x,y) - g(x+dx, y+dy) |$$

where Ω is a correlating region with a center at the point (x_0, y_0), for which we are seeking the disparity, f(x,y) and g(x,y) represent brightness in points of the base and scanned regions respectively, and $(dx, dy) \in P$

Generalizing calculation of correlation function from region $\mathbf{\Omega}$ to an arbitrary region $\boldsymbol{D}$:

$$\sigma(D, dx, dy) = \sum_{(x,y)\in D} | f(x,y) - g(x+dx, y+dy) |$$

(Then $\Re_3(x_0, y_0, dx, dy) = \sigma(\Omega, dx, dy)$.)

It is quite inefficient to calculate the correlation function for all N points independently. Indeed, if the correlating regions for two points overlap then computation of $\Re_3$ for each point involves the same summing over the overlapping region D so preliminary computation of $\sigma(D, dx, dy)$ and its use for calculation of the correlation function may be much more effective.

Two different methods exploiting this idea may be used for calculating the correlation function.

Ω is assumed being rectangular x_{corr} x $y_{corr.}$

Method A.

Let us divide left image into N_ω non-overlapping small regions ω_i; and for each region calculate the sum

If it is possible to represent the region Ω as a union of ω_i then we can calculate the correlation function as follows:

Let ω_i be squares of size (L_ω, L_ω). Two cases are possible.

1. Ω is a square with a side $L = N_L L_\omega$. The mesh granularity equals to the distance between centers of two adjacent correlating regions - L_ω.

2. Ω is a rectangle with sides $N_L L_\omega$ or $(N_L+1)L_\omega$ The mesh granularity equals to the distance between centers of two adjacent correlating regions - $L_\omega/2$.

To estimate the time of working of this algorithm as compared to the standard one let us calculate the number of elementary operations such as summing and subtraction. The standard algorithm involves $n_0=NL^2P$ such operations, where N is the number of points of a mesh, L is the linear size of the correlating region and P is the number of points in the scanned region.

The optimized algorithm consists of two steps. The number of operations in the first one (calculating $\sigma(\omega_i, dx, dy)$) is $n_1=N_\omega L_\omega^2 P$, and the number of operations in the second one (calculating $\Re_3(x, y, dx, dy)$) is $n_2=NN_L^2P$, where N_ω is the number of regions ω_i in an image and N_L^2 is the number of regions ω_i in Ω. Thus, the whole algorithm takes n = $(N_\omega L_\omega^2 + NN_L^2)P$ operations. The increase in time is

In the standard algorithm the correlating region of size 11x11 was used. Below are two examples of realization of the optimized algorithm in which the correlating region most closely approximates the square 11x11.

A1. ω_i are squares 4x4, correlating region is the square 12x12, mesh granularity is 4; $n_0/n=4.8$

A2. ω_i are squares 3x3, correlating region has the size 9x9, 9x12, 12x9 or 12x12, mesh granularity is 1.5; $n_0/n=8.3$

Thus, in the first case the increase in speed is 4.8 times and in the second 8.3 times.

Method B.

Let us denote $\Omega(x, y)$ as the square of size $(2\Delta+1)$x$(2\Delta+1)$ and $\omega(x, y)$ as the rectangle 1x$(2\Delta+1)$ with centers at (x, y):

We can write

(2)

,

(3)

Thus, the following algorithm can be used to calculate the values of the correlation function for the points of a rectangle $x_{min}\le x\le x_{max}$, $y_{min}\le y\le y_{max}$ with a correlation region $(2\Delta+1)$x$(2\Delta+1)$ and with a scanned region P:

- Calculate $\sigma(\omega(x, y), dx, dy)$ for the left border of a rectangle $x_{min}\le x\le x_{max}$, $y=y_{min}$ and for all (dx, dy) from P
- Calculate $\sigma(\omega(x, y), dx, dy)$ for all points of a rectangle $x_{min}\le x\le x_{max}$, $y_{min}\le y\le y_{max}$ using (2)
- Calculate $\sigma(\Omega(x, y), dx, dy)$ for the upper border of a rectangle $x=x_{min}$, $y_{min}\le y\le y_{max}$
- Calculate $\Re_3(x, y, dx, dy) = \sigma(\Omega(x, y), dx, dy)$ for all points of a rectangle $x_{min}\le x\le x_{max}$, $y_{min}\le y\le y_{max}$ using (3)

Assuming $x_{max} - x_{min} >> \Delta$, $y_{max} - y_{min} >> \Delta$ we can neglect the time of computation of the first and third steps. The second and the fourth steps require N uses of (2) and (3) accordingly, each formula uses 3 elementary operations. Thus the number of elementary operations can be estimated as n = 3NP. The increase in time is $n_0/n=L^2/3=40.3$ for a correlating region of 11x11.

Modifications of the algorithm for calculation of the correlation function.

These methods can only be used in the low level of the pyramidal algorithm since they require the scanned region to be the same for all points. However it is possible to modify them to deal with the high level where scanned regions for different points may differ.

Modification for the method A.

We will not use the preliminary step of computation of $\sigma(\omega_i, dx, dy)$. Instead we calculate $\sigma(\omega_i, dx, dy)$ only when necessity arises: suppose we want to calculate $\Re_3(x_0, y_0, dx, dy)$ and ω_i belongs to Ω; if $\sigma(\omega_i, dx, dy)$ has not been calculated before then calculate it and store the value, otherwise just use the value computed before.

Modification for the method B.

For the calculation of $\sigma(\omega(x, y), dx, dy)$ use the following algorithm:

- if $\sigma(\omega(x-1, y), dx, dy)$ has been calculated then use (2);

- if not use the direct summing.

Analogously for the calculation of $\sigma(\Omega(x, y), dx, dy)$ use the same algorithm:

- if $\sigma(\Omega(x, y-1), dx, dy)$ has been calculated then use (3);

- if not use the direct summing of $\sigma(\omega(x, y), dx, dy)$ whose values are calculated using the previous algorithm.

The points for which the correlation function is calculated should be processed in the following order: first the points in the first column from top to bottom, then the points in the second column, and so on.

The property of these modifications is that if used with the same scanned region for all points they involve the same number of calculations of auxiliary values ($\sigma(\omega_i, dx, dy)$ for the method A and $\sigma(\omega(x, y), dx, dy)$ for the method B) which are used as the original algorithms. However these modifications require additional overhead charges and thus run slower. The possible solution may be to run the original algorithm for some subset of P P' before applying the modified algorithm where

$$P' = \{(dx, dy) \in P_{\max} : h(P_{\max}, dx, dy) > \beta n\}$$

and β is determined experimentally.

The mesh granularity in the method *A1* is 4, in *A2* is 1.5 and in the method *B* is 1. Hence method *A1* may be used in the high level of the pyramidal algorithm and methods *A2* and *B* in the low level. Which method should be used depends on the required accuracy and time constraints: method *B* gives in 2.25 more points than method *A* but runs about 2 times slower.

3.5 Finding the maximum of the correlation function.

To determine the disparity we should find the maximum of the normalized correlation function on the scanned region *P*. The task is complicated by the fact that the function often has several local maxima on *P* so most methods for finding the maximum of a function do not work. We used the following method allowing to reduce the number of operations required for calculating correlation function up to four times as compared to the case of calculating the correlation function for all points of *P*.

• Set $P_0 = \{(dx, dy) \in P \mid \frac{dx}{2} \in N, \frac{dy}{2} \in N\}$

- points of *P* lying on the grid thinned two times

• (α) Calculate the correlation function for points of P_i

• Find the maximum of the correlation function $\Re_{3\max}$ among calculated values

• Set $P_{i+1} = P_i \cup \{(dx_0, dy_0) \in P \mid \exists (dx, dy) \in P_i :$

$| dx - dx_0 | \le 1, | dy - dy_0 | \le 1, \Re_3(dx, dy) \ge \Re_{3\max} - \Delta\Re_3\}$

• If $P_{i+1} \ne P_i$ then go back to (α)

The value of $\Delta\Re_3$ is determined experimentally from the condition that the number of points for which the minimum is found incorrectly is less than some small value say less than 1% from the total number of points.

This algorithm can be modified in such a way that for a certain number of points it will find disparity even if it does not lie in the scanned region practically without additional computational expenses. For that reason we replace the rule for choosing P_{i+1} with the following rule:

$$P_{i+1} = P_i \cup \{(dx_0, dy_0) \mid \exists (dx, dy) \in P_i :$$

$$| dx - dx_0 | \le 1, | dy - dy_0 | \le 1, \Re_3(dx, dy) \ge \Re_{3\max} -$$
$$- \Delta\Re_3(P, dx_0, dy_0)\}$$

where

$$\Delta\Re_3(P, dx_0, dy_0) = \begin{cases} \Delta\Re_3^{\,1}, (dx_0, dy_0) \in P \\ \Delta\Re_3^{\,2}, (dx_0, dy_0) \notin P \end{cases}$$

The disparity variations often do not exceed several pixels. It is a very rough scale for performing subsequent calculations and deciding about recognition. This problem is put away by introduction of *super-resolution*. Super-resolution is used to increase the granularity of scale by employing information from neighboring points. Simple and thus fast method of increasing resolution is used. Suppose the correlation function values are calculated in some scanning region. The maximum of the correlation is then found. If it lies on a boundary of the its coordinates are considered to be the sought disparity. Otherwise, the maximum and the nearest neighbors are interpolated by a convex surface (for example parabolic). The disparities then are calculated as coordinates of the surface crest. It is easy to see that the number of disparity gradations obtained by such method strongly exceeds the number of source points.

4. RESULTS AND CONCLUSION

The presented techniques were implemented in the framework of face recognition system [1] and were tested on a database of human faces. Figure 2 illustrates the source images. The white pixels on the figure represent some of correspondences found on images. Figure 3 displays disparity maps, calculated by the algorithm. The typical working time is 2-3 seconds on Pentium-II/300 processor.

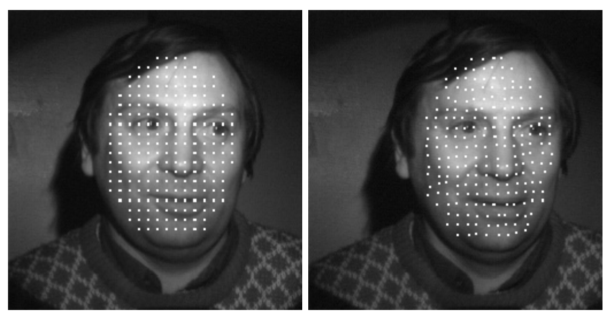

Figure 2 Source images

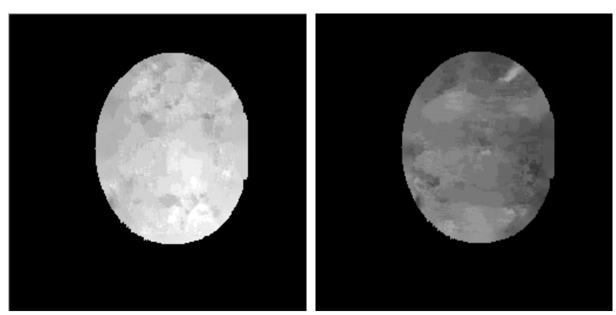

Figure 3 X- and Y-disparity maps

The testing of the described algorithms has shown that the precision of reconstruction is the same as in the stereoscopic system, while the time of execution slightly increases. It can be successfully used in various applications of monocular computer vision systems for reconstructing the 3D shape of object in a close to real time scale.

5. REFERENCES

[1] **Ivan A. Matveev, Alexander B. Murynin**. 3-D Surface Reconstruction in Automatic Recognition System. SPIE , 1999, Vol. 3516

[2] **Alexander B. Murynin, Ivan A. Matveev, Victor D. Kuznetsov**, "Automatic Stereoscopic System For Person Recognition", *//Proceedings of SPIE,* V.3516, 1998

[3] **Alexander B. Murynin, Victor D. Kuznetsov, Ivan A. Matveev** "Optimization of Informative Components for 3-D Object Recognition", *//Proceedings of SPIE,* V.3516, 1998.

[4] D.**Marr, T.Poggio** "A Computational Theory of Human Stereo-vision" //*Proceedings. Royal Society. London,* 1979, V.B204

[5] **U.R.Dhond, J.K.Aggarwal** "Structure from Stereo - A Review" //*IEEE Transactions on Systems, Man and Cybernetics* November/December 1989, V.19, N.6

[6] **B.K.P.Horn** "Height and Gradient from Shading" //*International Journal of Computer Vision,* 1990, V.5, N.1

[7] **Y.G.Leclerc, A.F.Bobick** "The Direct Computation of Height from Shading" //*IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1991

[8] **P.Fua, Y.G.Leclerc** "Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading" //*International Journal of Computer Vision,*1995, V.16

[9] **H.K.Nishihara** "Practical real-time imaging stereo matcher" //*Optical Engineering,* September/October 1984, V.23, N.5

[10] **R.Lengagne, J.P.Tarel, O.Monga** "From 2D Images to 3D Face Geometry" //*Proceedings of IEEE Second International Conference on Automatic Face and Gesture Recognition,* October 1996, Killington, USA

[11] **K.Hartt, M.Carlotto** "A Method for Shape from Shading using Multiple Images Acquired under Different Viewing and Lighting Conditions" //*IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1989

[12] **P.Fua** "Parallel stereo algorithm that produces dense depth maps and preserves image features" //*Machine Vision and Applications*, 1993, V.6

[13] **W.E.L.Grimson** "A Computational Implementation of a Theory of Human Stereo-vision" //*Phil. Transactions. Royal Society. London,* 1981, V.B292

Author(s):

Vladimir N. Kolmogorov, graduate student of Moscow Institute of Physics and Technology
Address: Kerchenskaya 1A1, Moscow, 113303, Russia
E-mail: vladimir.kolmogorov@usa.net

Ivan A. Matveev, postgraduate student of Moscow Institute of Physics and Technology.
Address: Kerchenskaya 1A1, Moscow, 113303, Russia
E-mail: matveev@ccas.ru

Alexander B. Murynin, Ph.D in Computer Science, Senior Researcher of Computing Center of RAS.
Address: Vavilova 40, Moscow, 119967, Russia
E-mail: murynin@ccas.ru