DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

Geometrically constrained sub-pixel disparity estimation from stereo images of the retinal fundus

Mohamad Kharboutly[1] ***, Carlos Vázquez***[1] ***, Stéphane Coulombe***[1] ***and Jacques de Guise***[1,2]
[1] ***École de technologie supérieure (ÉTS), Montréal, Québec, Canada***
[2] ***University of Montreal Hospital Research Centre (CRCHUM), Montréal, Québec, Canada***

Abstract

The aim of this study is to propose a precise disparity estimation process between retinal fundus stereoscopic images to assist during eye diagnosis and view generation for visualization. We propose a computer-vision-based solution that allows, from stereo images, the extraction of precise disparity information that could be used for clinical parameters estimation and/or the generation of multi-viewpoint images of retinal fundus. The proposed solution is based on a sub-pixel disparity estimation algorithm that takes into consideration the spherical shape of the retinal fundus. A sub-pixel approach is adopted in order to achieve high precision disparity estimation. Moreover, the a priori knowledge of the fundus geometric shape provides useful information for the regularization of the high precision disparity estimation process. Stereo images, with known ground truth, are used to evaluate the proposed algorithm and to demonstrate the high precision of estimated disparity in our methods.

Introduction

3D retinal surface visualization is becoming easy and handy with the current advances in image visualization technologies. Retinal image processing is a well-studied field and several studies about depth estimation of retinal fundus exists [1, 2, 3, 4].

The literature for the analysis of retinal images mainly deals with the issue of 2D images in order to facilitate the detection of lesions by ophthalmologists. However, accurate diagnosis often requires a 3D image of the retina. In particular, the analysis of the 3D shape of retinal fundus is essential for identifying lesions and estimating the extent of the lesion, and measuring eye pressure on the optic disc. Different methods are used to regenerate a 3D model of the fundus from stereo images. In [3], the authors estimate the epipolar geometry and projection matrices after a self-calibration, then they solve the correspondence problem to reconstruct the 3D fundus surface. In [7], the authors apply a plane+parallax algorithm to stereo images, which is followed by a mutual information-based disparity search stage. In [8], the authors use a multi-focusing technique to capture retinal fundus images and reconstruct its 3D surface. None of these studies assume that the retinal fundus can be approximated to a spherical surface.

In [5, 6] the authors assume a spherical shape for the eyeball. They use this information to help reconstruct the 3D shape of the fundus. One challenging step is to estimate the intermediate optical system parameters such as the contact enlarging lens focal distance. They model the mapping of this surface through the eye lens as a quadratic surface and perform camera calibration through matching correspondences on this quadratic surface. Their method is based on a two-stage optimization process. They estimate the quadratic surface using the correspondences and they minimize the error with respect to the camera poses by keeping the surface equation fixed. Secondly, they minimize the error with respect to the surface equation and camera poses to estimate the optical system parameters. This work aims to reconstruct the 3D shape of the retinal fundus, and does not use the estimated quadratic surface to increase the robustness of the disparity estimation process.

On the other hand, the authors in [10], state that the extraction of invariant geometric features is the first step not only for the 2D registration but also for the 3D reconstruction of retinal images. They proposed a robust method, using reliable landmarks, defined by the *Y-feature* , where vessels and nerves intersect. They used this method to extract robust corresponding points in order to rectify uncalibrated stereo images and then estimate the disparity without taking into consideration the quadratic surface constraint. However, for establishing disparities, the image correspondence problem must be solved throughout feature matching. Matching techniques, and thus constructing the disparity map, can be classified into two main categories. More specifically, techniques that construct the disparity map by solving the correspondence problem, pixel by pixel, are referred to as local, while techniques that consider the correspondence problem as a global optimization problem are referred to as global. Among the best-known techniques belonging into the mentioned categories are the differential matching, the cross correlation, graph cuts, the global energy optimization method and dynamic programming based methods to name a few. All local methods typically use an appropriate measure in order to quantify the existing similarity between the template window and the candidate one. Widely used similarity measures are the sum of squared differences (SSD), the sum of absolute differences (SAD), and the normalized cross-correlation (NCC) as well as their zero-mean counterparts. Among these measures, only the zero-mean normalized cross correlation is invariant to both shift and scale photometric distortion. This property is required in many stereo vision algorithms, especially in ophthalmic applications where the illumination of the scene is nonuniform. Another feature which is desired in a large number of applications is the ability of the matching algorithm for producing a disparity map with sub-pixel accuracy. Sub-pixel accuracy allows for a finer variation in depth levels, which otherwise are limited to values corresponding to integer disparities. This improves the precision of the depth estimation process and thus the quality of the synthesized images for multi-viewpoint visualization. In [9] the authors propose a new similarity measure which is based on the correlation coefficient called Enhanced Normalized

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

Cross Correlation (ENCC). More specifically, by using an appropriate linear interpolation scheme on the intensities of two adjacent candidate windows, a new similarity measure is introduced. This measure, although based on a linear interpolation scheme, does not demand the reconstruction of any intensity value, while at the same time it has infinite precision in its sub-pixel estimates.

In this paper, we propose a new algorithm for precise and robust sub-pixel disparity estimation exploiting the a priori knowledge of the retinal fundus shape. This method consists of, first, extracting reliable landmarks, defined by *Y-feature* points in stereo images. These points are then matched and used to generate a quadratic model of the retinal fundus. Secondly, a global sub-pixel disparity estimation process is developed under a geometric constraint defined by the quadratic surface.

The rest of the paper is organized as follows: Section 2 introduces the geometric model of the fundus surface and the its parameters estimation process. Section 3 describes the matching algorithm for sub-pixel disparity map. Section 4 discusses the algorithm implementation and its results. The last section concludes the paper and outlines future research directions.

Geometric constraint for fundus stereo images

In this paper we will consider the stereoscopic setup presented in figure 1. The two cameras are modeled as Pin-Hole in a canonical stereoscopic system. The stereo images are considered rectified, thus any point in the space will be mapped to corresponding image points through fundamental matrices $\mathscr{M}_l$ and $\mathscr{M}_r$ at the same vertical position.

Disparity geometric model

Assuming an orthogonal frame $O,\vec{x},\vec{y},\vec{z}$ and a retinal fundus (modeled as a sphere) with radius R, the coordinates (X,Y,Z) of a point M lying on the fundus surface should verify the sphere equation:

$$(X-C_x)^2+(Y-C_y)^2+(Z-C_z)^2=R^2, \tag{1}$$

where C_x, C_y and C_z are the coordinates of the sphere-shaped fundus center. The figure 2 shows the spherical shape and the optical path of a retinal fundus stereoscopic image capturing system. In [5] the author demonstrates that the point M is mapped through the optical systems into a point p which coordinates verify a quadratic equation of the form:

$$x^2+y^2+\alpha_z z^2+\alpha_{xz}xz+\alpha_{yz}yz+\alpha_x x+\alpha_y y+\alpha_z z+\alpha=0. \tag{2}$$

The projection of p to both camera plans Π_l and Π_r is computed using the cameras fundamental matrices $\mathscr{M}_l$ and $\mathscr{M}_r$. This will give two image points $m_l(u_l,v)=\mathscr{M}_l\times p^t$ and $m_r(u_r,v)=\mathscr{M}_r\times p^t$ in the left and right image coordinate system of the left and right cameras respectively. The disparity between m_l and m_r is defined by $d=u_l-u_r$. We can demonstrate that:

$$u_l=\frac{f}{z}(x+b/2),\ u_r=\frac{f}{z}(x-b/2),\ v=\frac{fy}{z}\ and\ d=\frac{fb}{z} \tag{3}$$

where f is the focal distance between the image plane and the camera center, and b is the distance between the two camera centers along the x axis. Performing simple mathematical manipulation, we can write:

$$x=\frac{b}{2}\frac{2u_l-d}{d},\ y=\frac{vb}{d}\ and\ z=\frac{fb}{d} \tag{4}$$

by replacing equation (4) into equation (2) and assuming $d\neq 0$ we can write:

$$G(u_l,v,d)= \tag{5}$$
$$u_l^2+v^2+a_1d^2+a_2u_ld+a_3vd+a_4u_l+a_5v+a_6d+a_7=0$$

where $(a_i, i=1...7)$ are real valued parameters.

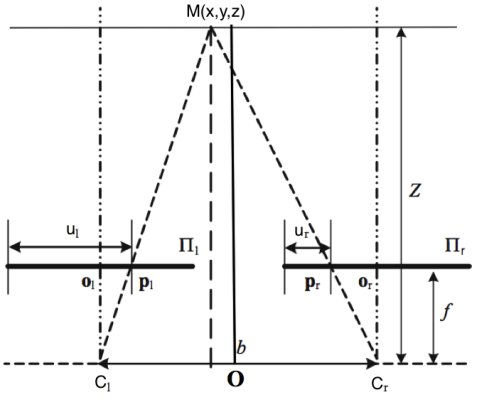


Figure 1: Stereoscopic canonical system

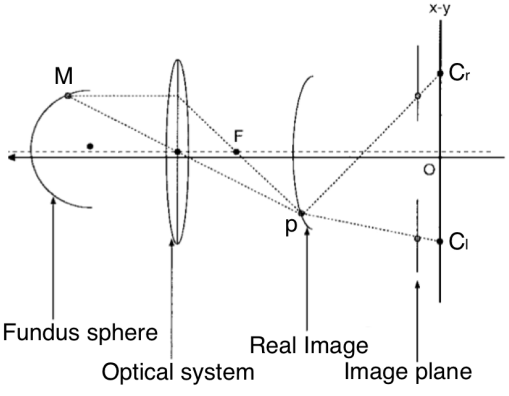


Figure 2: Stereoscopic setup for stereo retinal fundus image

The equation (5) presents a constraint between corresponding points in fundus stereo images that links the disparity value to the coordinates of the left point, considered as reference in this study. In order to integrate this constraint into the disparity estimation algorithm, the parameters of equation (5) should be estimated.

Identification of disparity geometric model parameters

In order to estimate the parameters in equation (5) a robust matching algorithm is used. As the model has 7 unknown parameters, at least 7 matching pairs of points are required to estimate these parameters. In [7], the author proposes a robust algorithm

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

to detect special features in the retinal fundus image. This algorithm is based on the detection of intersection points between blood vessels called *Y-feature* points. The position of a *Y-feature* has at most 3 strong responses from different directional Laplacian of Gaussian filter outputs. Basically, the algorithm generates a list of potential points using 6 directional Laplacian filters. The obtained filtered outputs are analyzed using the Principal Components Analysis (PCA) method for every pixel. Image locations having at least three large component values are considered as seed points in order to guarantee that at least three gradient directions are present in the vicinity of the *Y-feature* .

The *Y-feature* model has 8 degrees of freedom (DOF), $\mathbf{Y} = (u, v, \theta_1, \theta_2, \theta_3, w_1, w_2, w_3)$ which includes the center position (u, v), the orientation angles of the three branches attached to the center position $\theta_1, \theta_2, \theta_3$, and three widths for each branch w_1, w_2, w_3. The length of each branch, L, is fixed. Using the geometric properties of the *Y-feature* in retinal images, the arms are constrained not to be too close or far away from each other. The width is also limited to be between the minimum and maximum size of the vessels to be detected. Given the initial position of the *Y-feature* provided by the estimated seed points, we fit the articulated model using a gradient descent method, minimizing the following energy:

$$F(\mathbf{Y}) = \frac{1}{2}\sum_{i=1}^{3}\int_{-w_i}^{w_i}\int_{0}^{L}[(-1)^m \tilde{\mathscr{I}}(u_i, v_i)^2 + \tilde{\mathscr{G}}(u_i, v_i)^2]dwdl \tag{6}$$

where u_i, v_i are the coordinates of a point in the articulated *Y-feature* model, $u_i = u + w\sin(\theta_i) + l\cos(\theta_i)$ and $v_i = v + w\cos(\theta_i) + l\sin(\theta_i)$ where (u, v) is the center position of the model and i is the index of the considered arm. $\tilde{\mathscr{I}}(u_i, v_i)$ is the interpolated value of intensity image $\mathscr{I}$ at the point u_i, v_i and $\tilde{\mathscr{G}}(u_i, v_i)$ is the gradient value of the interpolated intensity image $\tilde{\mathscr{I}}$ at the point u_i, v_i. The parameter $m = 0$ to find dark vessel, and $m = 1$ to detect bright vessels. Additionally, the author constrains the angles $\theta(\theta_1, \theta_2, \theta_3)$ and widths $w(w_1, w_2, w_3)$ of the branch to be within a specified range of values θ_{min} and θ_{max} for angles and w_{min} and w_{max} for widths. Enforcing these constraints could be achieved using Lagrange multiplier for solving the constrained optimization problem. The author uses a penalization approach based on the use of a barrier function B for enforcing the inequality constraints:

$$B(\theta, w) = \sum_{j=1}^{3}\left(\frac{1}{(\theta_i - \theta_{min})(\theta_{max} - \theta_i)} + \frac{1}{(w_j - w_{min})(w_{max} - w_j)}\right)$$

The extraction of *Y-feature* in the image consists of initializing the model using feature point location and orientations and fitting the articulated *Y-feature* to the image features by minimizing the function:

$$E(\mathbf{Y}) = F(\mathbf{Y}) + (1 - \beta)B(\theta, \mathbf{w}) \tag{7}$$

where β is the trade-off between the goodness of fit to image features and the constraints on the orientation of the arms and their thickness. The function E is minimized iteratively using a gradient-based approach. For more detailed information refer to [11].

In order to use this approach on stereo images to select matching points, a matching strategy should be adopted. As we use rectified images, the corresponding matched points have the same vertical position. This will limit the search to only one direction. The search begins from a minimum disparity value. Using traditional matching algorithms (cross-correlation), a first estimate will be detected. Then, this point will be considered as a seed point in the right image and a new energy optimization process (equation (6)) will be executed to find a more precise position of the matched points. This energy optimization will be constrained with a fixed vertical position ($v_i = \text{constant}$).

A first estimate of fundus model parameters is computed using at least 7 *Y-feature* matched points. If more than 7 points were found, the best estimate is computed using the least-square minimization approach with the form

$$\tilde{P} = \arg\min_{P}(\|B + A \times P\|_2^2) \tag{8}$$

where

$$P = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \end{bmatrix}^t \tag{9}$$

A is the matrix and B the vector with k^{th} row respectively equal to:

$$A_k = \begin{bmatrix} d_k^2 & u_{l,k}d_k & v_k d_k & u_{l,k} & v_k & d_k & 1 \end{bmatrix} \tag{10}$$

$$B_k = \begin{bmatrix} u_{l,k}^2 + v_k^2 \end{bmatrix} \tag{11}$$

and k goes from 1 to S being S the number of *Y-feature* matched points.

Geometrically constrained sub-pixel disparity estimation

The most commonly used sub-pixel approach in stereo matching is based on polynomial interpolation, including correlation interpolation methods and intensity interpolation methods. In [9] the author proposes a similarity measure based on the correlation coefficient that is called Enhanced Normalized Cross Correlation (ENCC). This measure is based on a linear interpolation scheme. It does not require the reconstruction of any intensity value, while at the same time it has infinite precision in its sub-pixel estimates. This algorithm is adopted in this paper.

Let us consider a rectified stereo image, with $\mathscr{I}_l(i, j)$ and $\mathscr{I}_r(i, j)$ denoting their intensity functions. The width and height of both images are the same and noted L_w and L_h respectively. The stereo correspondence problem aims to find a non-negative disparity map $D(i, j)$ such that the following relation approximately holds

$$\mathscr{I}_l(i, j) \approxeq \mathscr{I}_r(i - D(i, j), j) \tag{12}$$

In order to solve the image correspondence problem in a local window-based method, let us consider that $W(u, v)$ denotes an image window of size $N1 \times N2$ with its center located at the point with coordinates u, v, and let

$$w(u, v) = \begin{bmatrix} w_1 & w_2 & \dots & w_{N-1} & w_N \end{bmatrix}^t \tag{13}$$

be the vector resulting by stacking up the columns of the window $W(u, v)$, where $N = N1N2$ is its length. Let us also define the zero

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

mean normalized version of vector $w(u,v)$ as

$$\bar{w}^{\circ}(u,v) = \frac{w(u,v) - \bar{w}(u,v)}{\|w(u,v) - \bar{w}(u,v)\|_2} \tag{14}$$

where $\bar{w}(u,v)$ and $\|w(u,v)\|_2$ denote its mean value and Euclidean norm respectively. By selecting a template window $W_l(u,v)$ in the reference image, and a window $W_r(u-d,v)$ in the matching image, we can define, using the above notation, their correlation coefficient as the inner product of the vectors

$$\rho_{u,v,d} = \bar{w}_l^{\circ t}(u_l,v)\bar{w}_r^{\circ}(u_r-d,v) \tag{15}$$

and use it as a similarity measure for the centers of the above defined windows.

The correlation coefficient of equation (15) can be generalized to become a function of the continuous spatial variable, τ, $\rho_{u,v,d}(\tau)$. Let us therefore introduce the following $N-$dimensional vector function:

$$w(u+\tau,v) = w(u,v) + (1-\tau)(w(u-1,v) - w(u,v)) \tag{16}$$

which is a continuous linear function of the spatial variable $\tau \in [0,1]$. Notice that if $\tau = 0$, $w(u,v,\tau) = w(u-1,v)$ and if $\tau = 1$, $w(u,v,\tau) = w(u,v)$.

The goal now is to incorporate the intensity vector function of τ into the similarity measure defined in equation (15). To this end, let us define the following correlation function:

$$\rho_{u_l,v,d}(\tau) = \bar{w}_r^{\circ t}(u_l - d - \tau, v)\bar{w}_l^{\circ}(u_l,v) \tag{17}$$

Using the definitions of the inner product and zero mean normalized vector of equation (15), and after some mathematical manipulations, equation (17) can be rewritten as :

$$\rho_d(\tau) = \frac{\rho_d + \tau(\rho_d - \lambda\rho_{d-1})}{\sqrt{(1+\lambda^2 - 2\lambda r)\tau^2 + 2(1-\lambda r)\tau + 1}} \tag{18}$$

where

$$\lambda = \frac{\|w_r(u_l - d - 1, v) - \bar{w}(u_l - d - 1, v)\|_2}{\|w_r(u_l - d, v) - \bar{w}(u_r - d, v)\|_2} \tag{19}$$

is the ratio of norms of the adjacent windows and

$$r = \bar{w}_r^{\circ t}(u_l - d, v)\bar{w}_r^{\circ}(u_r - d - 1, v) \tag{20}$$

their correlation coefficient.

Geometric constraint disparity estimation

After introducing the geometric model of the fundus surface and a sub-pixel accuracy similarity measurement operator, this section will discuss how the depth map is estimated under a geometric constraint. The proposed algorithm is based on an energy maximization function. This energy will evaluate the current disparity estimation with respect to two different criteria. The energy function is:

$$E(D,P) = (1-\alpha)E_{Data}(D) + \alpha E_{Geo}(D,P) \tag{21}$$

where D is the estimated disparity map, P is the parameters vector of the quadratic surface defined in equation (9) and α is a weighting parameter. The goal is to find $\tilde{D}$ that maximizes $E(D,P)$:

$$\tilde{D} = \arg\max_D \{E(D,P)\} \tag{22}$$

Data energy

The data energy E_{Data} is the energy that measures the similarity between all windows $w_l(u_l,v)$ and $w_r(u_l-d,v)$ where d is the estimated sub-pixel disparity $D(u_l,v)$. The data energy can then be computed by summing all the correlation values $\rho(u_l,v,d)$ at all pixels. Thus:

$$E_{Data}(D) = \sum_{u_l=1}^{L_w}\sum_{v=1}^{L_h} \rho(u_l,v,D(u_l,v)) \tag{23}$$

where $D(u_l,v)$ is the disparity value at the point (u_l,v). The maximum value of the correlation function at a given pixel, is equal to unity. Thus, the maximum data energy value is equal to L_hL_w.

Geometric energy

$E_{Geo}(D,P)$ is the energy that measures how much the sub-pixel disparity verifies the geometric constraint. This energy is computed using the geometric model presented in equation (5). The function $G(u,v,d)$ from equation (5) is a quadratic surface that represents a relation between a point (u,v) and its corresponding $(u-d,v)$. More the point (u,v,d) is close to the surface more $G(u,v,d)$ is close to zero. Thus, the geometric energy can be defined as follows:

$$e_{Geo}(u_l,v,D,P) = \frac{1}{\sqrt{1+G(u_l,v,D(u_l,v))^2}} \tag{24}$$

where P is the model parameter defined in equation (9) and used in the function $G(ul,v,D(u_l,v))$. It is obvious that the value of $e_{Geo} \in [0,1]$. It is equal to 1 when $G(u_l,v,D(u_l,v))$ is zero which means the point $(u_l,v,D(u_l,v))$ is on the surface. It decreases when $G(u_l,v,D(u_l,v))$ increases. By doing so, the two energies are compatible and can be added together.

The total geometric energy can be computed as follows:

$$E_{Geo}(D,P) = \sum_{u_l=1}^{L_w}\sum_{v=1}^{L_h} e_{Geo}(u_l,v,D,P). \tag{25}$$

Gradient descent solver

The final disparity map is the one that maximizes the energy function (21). One way to solve the optimization problem is to use the gradient-descent method. This method requires the computation of the energy gradient with respect to the disparity values. By slightly varying the disparity value d at each pixel (u_l,v) in D a gradient vector is computed:

$$\frac{\partial E(D,P)}{\partial D} \approxeq \frac{E(D+\Delta,P) - E(D-\Delta,P)}{2\Delta} \tag{26}$$

If the energy is to be maximized then:

$$D_{new} = D_{old} + \gamma\Delta E(D,P) \tag{27}$$

where Δ and γ are internal parameters that affect the convergence speed and stability of the algorithm.

Algorithm

The criteria optimization process is essentially based on the gradient descent method. The geometric constrained sub-pixel disparity estimation algorithm is described as following.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

- Step 1 find list of corresponding points using *Y-feature* matching algorithm,
- Step 2 estimate Fundus geometric model parameters P using the *Y-feature* matched list,
- Step 3 compute the total energy according to, (21), (23) and (25)
- Step 4 apply the gradient descent algorithm to compute the next disparity map D and
- repeat step 3 and 4 until convergence.

Simulation and Experimental results

Stereo images and ground truth generation

For simplicity, we will consider that the stereo images are the projection of a spherical surface rather than a quadratic surface. In fact by neglecting the optical system, the retinal fundus, considered as a spherical surface, is directly projected onto the cameras image planes Π_l and Π_r. This supposition will only affect the estimated values of the geometric model parameters (5). A 3D modeling software is used to create a 3D spherical model. In this study the software *Blender* is used. Under Blender, an unwrapping function will allow to project a 2D image onto a 3D spherical surface (see fig. 3a). Note that this process may seem to be unreasonable, but the main objective here is to test the proposed method and not to reconstruct a real 3D surface. Once the 3D spherical model is created, we define two different positions of the camera centers, noted $C_l = [C_{lx}, C_{ly}, C_{lz}]$ and $C_r = [C_{rx}, C_{ry}, C_{rz}]$. C_l and C_r are slightly displaced along the x axis with known distance b (see fig. 1). The displacement is computed in a way that will generate stereo images with a large interval of disparity values. All parameters, including the sphere radius, its center position, the camera centers position and its focal distance are adjusted and well-defined. The intrinsic parameters of the camera are the same at both positions and they are presented in the intrinsic matrix $\mathscr{M}^{int}$. The extrinsic parameters include the coordinate of the camera centers in the world frame and they are presented in $\mathscr{M}_l^{ext}$ and $\mathscr{M}_r^{ext}$. A point $M(x,y,z)$ in the world frame is mapped into $p_l(x,y,z)$ and $p_r(x,y,z)$in the left and right camera frame through the matrices $\mathscr{M}_l^{ext}$ and $\mathscr{M}_r^{ext}$ respectively. The projection of points p_l and p_r onto the image plane is computed through the matrix $\mathscr{M}^{int}$. Thus, the image of $p_r(x,y,z)$ is $m_r(u_r,v)$ on the right camera and the image of $p_l(x,y,z)$ is $m_l(u_l,v)$ on the left camera plane (see figs. 4a and 4b).

$$m_l = \begin{bmatrix} u_l \\ v \\ 1 \end{bmatrix} = \frac{1}{z_{c,l}} \mathscr{M}^{int} \times \mathscr{M}_l^{ext} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{28}$$

$$m_r = \begin{bmatrix} u_r \\ v \\ 1 \end{bmatrix} = \frac{1}{z_{c,r}} \mathscr{M}^{int} \times \mathscr{M}_r^{ext} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{29}$$

where

$$\mathscr{M}^{int} = \begin{bmatrix} \frac{f}{h_x} & 0 & O_x \\ 0 & \frac{f}{h_y} & O_y \\ 0 & 0 & 1 \end{bmatrix}, \; z_{c,l} = z - C_{lz}, \; z_{c,r} = z - C_{rz} \tag{30}$$

and O_x, O_y are the image center coordinates in pixels, h_x and h_y are pixel dimension in *meter* and f is the focal distance of the camera (see fig. 1) and

$$\mathscr{M}_l^{ext} = \begin{bmatrix} I_3 & C_l^t \\ 0 & 1 \end{bmatrix}, \mathscr{M}_r^{ext} = \begin{bmatrix} I_3 & C_r^t \\ 0 & 1 \end{bmatrix} \tag{31}$$

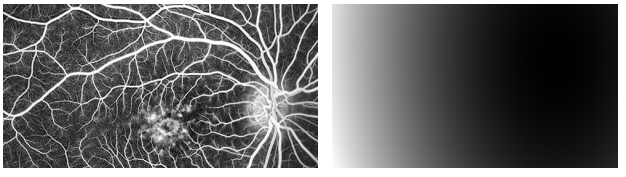
where I_3 is 3×3 identity matrix.

A point $m(u,v)$ can be mapped back to $M(x,y,z)$ by computing the line of sight using $\mathscr{M}_{int}^{-1}$ and the intersection with the fundus surface defined in equation (1).

The ground truth image is computed using $\mathscr{M}_l^{ext}$, $\mathscr{M}_r^{ext}$ and $\mathscr{M}^{int}$. The ground truth image is the disparity value at each pixel presented in a gray-scale image (see fig. 3b). The disparity d is equal to:

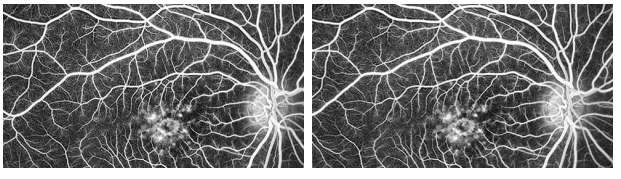
$$d = u_l - u_r. \tag{32}$$

u_l and u_r are the x-coordinate of points m_l and m_r respectively.



(a) The retinal fundus reference image (b) The ground truth of the retinal spherical model

Figure 3: This image was originally published in the ASRS Retina Image Bank. Author: James B. Soque. Photographer: James B. Soque. Title: Stereo View of Retinal Angiomatous Proliferation in Age-Related Macular Degeneration. Retina Image Bank. 2012; Image Number: 852. ©The American Society of Retina Specialists



(a) Left image (b) Right Image

Figure 4: Stereo images of retinal fundus

Algorithm implementation

Step 1 Y-Feature extraction

By applying the *Y-feature* extraction algorithm described previously, a list of *Y-feature* centers is generated and presented in the figure 5a. The parameters of the *Y-feature* algorithm are adapted to the vessels size inside the image. On the other hand, the matching process of *Y-feature* points is performed in two steps. First a local matching algorithm is used. The resulting matched point is then used as initial position in the *Y-feature* optimization in the right image. In this case, the optimization of criteria (6) is constrained to be only along the x-axis, thus the $y-coordinate$ value is considered fix. At least 7 matched pairs are required to estimate the surface parameters equation (5). The figure 5a and 5b show 10 matched points, in the left and right images, generated by the *Y-feature* algorithm, referred by the cross and square symbols respectively.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

Step 2 Model parameters estimation

Using the least-square solver, the parameters of equation (5) are estimated. First we construct the matrix described in (10) and (11). In this example, a list of 10 points are matched as presented in 5a. These points are used to create the 10 by 7 matrix A. Solving the least-square minimization problem $B + A \times P = 0$ where P gives the model parameters $\{a_i\}_{i=1...7}$.

Step 3 Energy computation

In this step, we calculate the energy of the current disparity map. This energy is formed by adding two different energy functions (equation (21)). The first one consists of a data energy which measures the similarity with respect to the disparity D and the second one consists of a geometric energy which measures how close the disparity D is to the geometric model defined by parameters vector P. A first estimate of the disparity map is required to compute this energy. Thus, we choose to start by the disparity map computed from the geometric model directly:

$$D_0(u,v) = \min_{\mathrm{d}}(u^2 + v^2 + a_1 d^2 + a_2 ud + a_3 vd + a_4 u + a_5 v + a_6 d + a7) \tag{33}$$

The energy at the iteration i is computed with respect to equation (21).

$$E_i(D_i,P) = (1-\alpha)E_{Data,i}(D_i) - \alpha E_{Geo,i}(D_i,P) \tag{34}$$

where

$$E_{Data,i}(D_i) = \sum_{v=1}^{L_h}\sum_{u=1}^{L_w} \rho(u,v,D_i(u,v)) \tag{35}$$

and

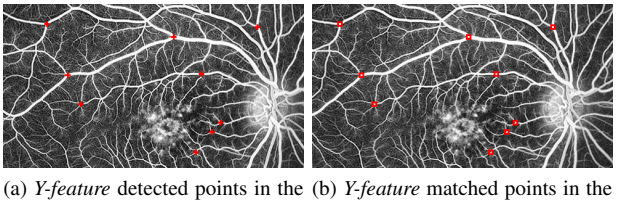
$$E_{Geo,i}(D_i,P) = \sum_{v=1}^{L_h}\sum_{u=1}^{L_w} e(u,v,D_i(u,v),P) \tag{36}$$

Step 4 Optimal disparity Estimation

After computing the energy of the initial estimate of the disparity map, we use the gradient descent method described previously in equation (26) to converge towards the optimal solution. From $E_i(D_i)$ we compute the new estimate D_{i+1} of D_i:

$$\frac{\partial E_i(D_i,P)}{\partial D_i} = \frac{E_i(D_i+\Delta,P) - E_i(D_i-\Delta,P)}{2\Delta} \tag{37}$$

$$D_{i+1} = D_i + \gamma \frac{\partial E_i(D_i,P)}{\partial D_i} \tag{38}$$



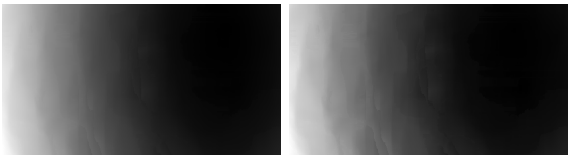
(a) *Y-feature* detected points in the left image (b) *Y-feature* matched points in the right image

Figure 5: *Y-feature* matched points in the left and right images.

Steps 3 and 4 are repeated until the algorithm converges. The convergence state is defined when the energy difference between two consecutive iterations is smaller than 1%:

$$Err_{en,i} = \left| \frac{E_{i+1} - E_i}{E_i} \right| \tag{39}$$

The figure 6a shows the estimated disparity map for $\alpha = 0.3$, $\Delta = 0.004$ and $\gamma = 0.5$.



(a) The resulting estimated disparity map for $\alpha = 0.3$ (b) The estimated disparity map without geometric constraints ($\alpha = 0$)

Figure 6: (a) Estimated disparity map for $\alpha = 0.3$. (b) Disparity map without geometric constraints ($\alpha = 0$)

A general approach to evaluate the performance of the algorithm is to compute the root-mean squared error with respect to the ground truth [12]:

$$R = \frac{\left(\sum_{u=1}^{L_w}\sum_{v=1}^{L_h} |D(u,v) - D_{GT}(u,v)|^2\right)^{\frac{1}{2}}}{\sum_{u=1}^{L_w}\sum_{v=1}^{L_h} |D_{GT}(u,v)|} \tag{40}$$

where $D(u,v)$ is the estimated disparity and $D_{GT}(u,v)$ is the ground truth data. In this example, where $\alpha = 0.3$ the R value is equal to 0.4%, while without the geometric constraint, the R value is equal to 0.8%.

Experimentation on real stereo images

An original way to evaluate this method is to apply it on real stereo images of the retinal fundus. Fundus images, at high resolution, can be found in the retinal image bank (http://imagebank.asrs.org/). In order to evaluate the estimated disparity using our method, some ground truth data is required. One way to generate ground truth data is to manually extract corresponding points at full image resolution. This procedure provides sub-pixel ground truth data after reducing the image resolution. In this example, we use the stereo images presented in figures 7a and 7b. Each image has a full resolution of 800×900 pixels and the scaling factor is considered to be 0.5. The figures 7a and 7b also show the manually extracted matching points.

In the following, a reduced version of the images in 7a and 7b are used as well as the extracted and scaled ground truth data.

First, we start by applying the *Y-feature* extraction algorithm. The purpose of this algorithm is to extract robust corresponding points that will be used in the quadratic surface estimation. The figures 8a and 8b show 10 extracted *Y-feature* points in both images.

Next, the model parameters, described by the vector P, are estimated using the least-square solver applied to the matrix A (equation (10)) and vector B (equation (11))

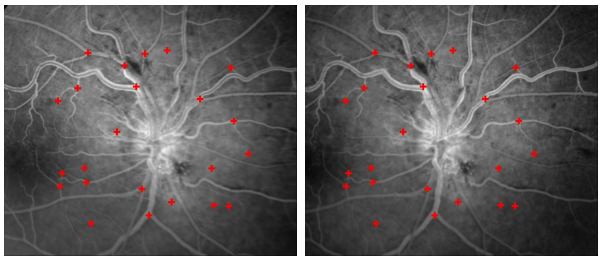
Finally, the optimal estimated disparity map is computed by optimizing the energy function (see equation (22)).

Figure 9 shows the RMS error between the ground truth matched points and their estimated disparity for 1-unconstrained

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

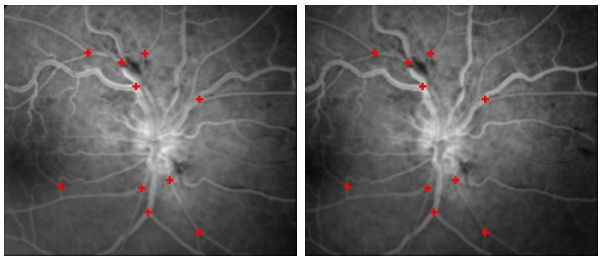
disparity estimation ($\alpha = 0$), 2- fully constraint disparity estimation ($\alpha = 1$) and 3- an intermediate value ($\alpha = 0.3$). The ground truth matched points are sorted and divided into 3 categories. The first category consists of points with textured neighborhood (e.g. vessels intersection) and close to *Y-feature* generated points. The second category consists of other points having textured neighborhood. The last category consists of points with low textured neighborhood. The figure 9 shows that the points of the first category have low RMS error ($< 4\%$) in the two extreme cases, $\alpha = 0$ and $\alpha = 1$. On the other hand, the points lying in the second category have low RMS error for $\alpha = 0$, and bigger RMS error (7%) when $\alpha = 1$. Finally, the points lying in the third category have RMS error (13%) at $\alpha = 0$ bigger than RMS error 5% at $\alpha = 1$.

This experimental result explains the fact that the data energy has a bigger contribution in the total energy at points with textured neighborhood, while the geometric energy contributes significantly more in low textured regions. Thus, by choosing an appropriate value for the parameter α, or by varying its value with respect to the texture of the region, we expect better results. In the figure 9, we can see that for $\alpha = 0.3$, we obtain smaller RMS errors compared to $\alpha = 0$ in the region with low texture (category 3). With this value of α, a small increase of the RMS error is observed in points of the second category which is expected and not considered critical as its value is less than 4%. In real fundus stereo images, the quadratic surface assumption is only an approximation, and the geometric model should only be considered for regularization of the estimated disparity. Thus, it is advised to consider small values of the parameter α (e.g. $\alpha < 0.5$).



(a) The left image with manually selected points (b) The right image with manually selected points

Figure 7: This image was originally published in the ASRS Retina Image Bank. Author: James B. Soque. Photographer: James B. Soque. Title: Diabetic Retinopathy Optic Nerve Edema, Fluorescein Angiogram, Stereo. Retina Image Bank. 2015; Image Number: 24867. ©The American Society of Retina Specialists



(a) *Y-feature* detected points in the left image (b) *Y-feature* matched points in the right image

Figure 8: *Y-feature* matched points in the left and right images.

Conclusion

In this paper we proposed a sub-pixel disparity estimation algorithms that takes into consideration the geometric shape of the retinal fundus. First, we detect the *Y-feature* points with their correspondences in stereo images. Then we generate a quadratic surface that models the retinal fundus using *Y-feature* matched pair points. The quadratic model is used to enhance the accuracy of the sub-pixel disparity estimation process. The simulated results have shown an improved precision of the geometrically constrained disparity map by comparing it to the unconstrained sub-pixel disparity estimation process. We are currently exploring the interaction between image characteristics and the regularization factor in order to find the optimal value.

Acknowledgments

Authors would like to acknowledge the help of Diego Joshua Martinez Pineda in implementing the first version of the *Y-feature* detection and matching algorithm. This work was partially supported by MITACS throughout its Globalink program and its Accelerate program in collaboration with the Institut de l'Œil des Laurentides (IOL).

References

[1] Ataer-Cansizoglu, E., Taguchi, Y., Kalpathy-Cramer, J., Chiang, M.F., and Erdogmus, D., Analysis of shape assumptions in 3D reconstruction of retina from multiple fundus images. In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, pp. 1502-1505. (2015).

[2] Hernandez-Matas, C., Zabulis, X., and Argyros, A. A., Retinal image registration based on keypoint correspondences, spherical eye modeling and camera pose estimation. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, pp. 5650-5654. (2015).

[3] Bansal, M., Sizintsev, M., Eledath, J., Sawhney, H., Pearson, D. J. and Stone, R. A., 3D optic disc reconstruction via a global fundus stereo algorithm. In Engineering in Medicine and Biology Society

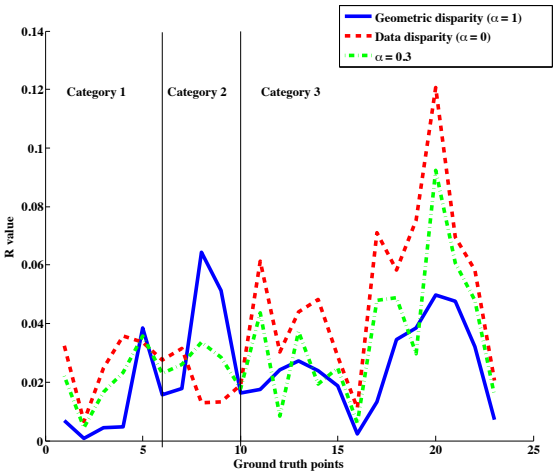


Figure 9: RMS error computed for $\alpha = 0$, $\alpha = 0.3$ and $\alpha = 1$ at the ground-truth manually extracted points.

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-034

(EMBC), 2013 35th Annual International Conference of the IEEE, pp. 5877-5882. (2013).

[4] Martinez-Perez, M. E., and Espinosa-Romero, A., Three-dimensional reconstruction of blood vessels extracted from retinal fundus images. Optics express, 20(10), 11451-11465. (2012).

[5] Deguchi, K., Noami, J., and Hontani, H., 3D fundus pattern reconstruction and display from multiple fundus images. In Pattern Recognition, Proceedings of the 15th International Conference on, Vol. 4, pp. 94-97. (2000).

[6] Deguchi, K., Kawamata, D., Mizutani, K., Hontani, H., Wakabayashi, K., 3D fundus shape reconstruction and display from stereo fundus images. IEICE Transactions on Information and Systems, 83(7), 1408-1414.(2000).

[7] Choe, T. E., Cohen, I., Medioni, G., Walsh, A. C., and Sadda, S. R., Evaluation of 3D shape reconstruction of retinal fundus. In Medical Image Computing and Computer-Assisted InterventionMICCAI, pp. 134-141. (2006).

[8] Motta, D., de Matos, L., de Souza, A. C., Marcato, R., Paiva, A. and de Carvalho, L. A. V., All-in-focus imaging technique used to improve 3D retinal fundus image reconstruction. In Proceedings of the 30th Annual ACM Symposium on Applied Computing, pp. 26-31. (2015).

[9] Psarakis, E. Z., and Evangelidis, G. D., An enhanced correlation-based method for stereo correspondence with subpixel accuracy. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1, pp. 907-912. (2005).

[10] Cheriyan, J., Menon, H. P., and Narayanankutty, K. A., 3D Reconstruction of Human Retina from Fundus ImageA Survey. International journal on modern engineering research, volume2, pages 3089-3092.(2012).

[11] Choe, T. E., and Cohen, I., Registration of multimodal fluorescein images sequence of the retina. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1, pp. 106-113.(2005).

[12] Scharstein, D., and Szeliski, R., A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision, 47(1-3), 7-42. (2002).

Author Biography

Mohamad Kharboutly received his BS in Electrical engineering from the Lebanese University (2007), and his PhD in control systems from the university of franche-comté, in France (2010). He has worked in the control system department of FEMTO-ST, France, for two years and currently he is working in the software engineering department at ETS, Montreal, Canada. His work has focused on the development of 3D disparity estimation algorithms.

Stéphane Coulombe received the B.Eng. degree in electrical engineering from École Polytechnique de Montréal, Canada, in 1991, and the Ph.D. degree from INRS-Telecommunications, Montréal, in 1996. He is currently a Professor in the Software and IT Engineering Department, École de technologie supérieure (ÉTS). From 1997 to 1999, he was with Nortel in the Wireless Network Group, Montréal. From 1999 to 2004, he was with the Nokia Research Center, Dallas, TX, as a Senior Engineer and as a Program Manager in the Audiovisual Systems Laboratory. He joined ÉTS in 2004, where he currently carries out research and development on video processing and systems.

Carlos Vázquez received the B. Eng. and M.Sc. degrees in 1992 and 1997 respectively from the ISPJAE technical university in Havana, Cuba. He received his Ph.D. degree in Telecommunications from the INRS-EMT, Montréal, Canada, in 2003. He was a Research Scientist with the Communications Research Centre Canada (CRC) from 2005 to 2013 when he joined the ÉTS Montréal. Dr. Vázquez is currently associate professor at the software and IT engineering department of the ÉTS. He is currently involved in 3D-TV, 3D vision systems, Multi-View video and 3D object extraction research.

Jacques A. de Guise is Full Professor at the Department of Automated Production Engineering of École de technologie supérieure (ÉTS). After completing his studies in Electrical Engineering at École Polytechnique Montréal in 1977, he did his graduate studies at the Biomedical Engineering Institute of École Polytechnique and the University of Montréal, where he obtained his PhD in 1984. He is also Associate Professor at the Department of Surgery, Faculty of Medicine, University of Montréal and an active member of the Centre Hospitalier de l'Université de Montéral Research Center (CRCHUM). He is the Director of the Imaging and Orthopaedics Research Laboratory (LIO) of ÉTS and CRCHUM. He holds the Canada Research Chair in 3D Imaging and Biomedical Engineering and the Marie-Lou and Yves Cotrel Research Chair in Orthopaedics at the Department of Surgery, Faculty of Medicine, University of Montréal. He is Fellow of the Canadian Academy of Engineering since 2011.