A Stochastic Optimization to Dense Stereo Matching*

M. Ouali[1] — K. Bouazza[1] — H. Lange[2]

[1]*Département d'Informatique,*
Université d'Oran Es-Senia, Oran, Algérie
[2]*Image Analysis R&D Group,*
Aperio Technologies, Vista, San Diego, CA, USA

Abstract. *This work presents a general system that achieves an energy minimization-based dense stereo matching through simulated annealing. Dense stereo matching is based on point matching. We show the performance of our approach compared to correlation. We use an optimization method in order to take into consideration the global aspect of the problem, as opposed to the correlation that acts locally on windows, and be able to make this module cooperate with other early vision modules, for instance shape from shading and photometric stereo. The stereo matching problem is an ill-posed problem where the global minimum is hidden by local minima and where the notion of gradient does not exist. For this reason, the simulated annealing algorithm seems the most suitable to solve the stereo matching problem. The constraints of the stereo matching are expressed as an energy functional and elementary transformation.*

Résumé : *Ce travail présente un système de mise en correspondance d'images stéréoscopiques à base d'optimisation stochastique, la mise en correspondance des images stéréoscopiques utilise principalement les primitives de ype points. Nous montrons les performances de notre approche par rapport à la corrélation. Nous avons exprimé le problème de mise en correspondance sous forme optimisationnelle pour mettre en avant son aspect global, alors que la corrélation agit localement. L'optimisation permet aussi de faire coopérer ce module avec d'autres modules de vision de bas niveau tels que la forme à partir de l'ombrage et la stéréo-photométrie. De plus, le problème de mise en correspondance est un problème mal-posé, où le minimum global est caché par des minima locaux et où la notion de gradient n'existe pas. Pour cette raison, l'algorithme du recuit simulé semble le plus adapté. Les contraintes de mise en correspondance sont exprimées par une fonctionnelle d'énergie et des transformations élémentaires.*

Keyword : *Stereo disparity, stochastic optimization, image matching, constraints, resemblance, epipolar, uniqueness, continuity*

Mots clés : *Disparité, optimisation stochastique, appariement d'images, contraintes stéréoscopiques, ressemblance, épipolaire, unicité, continuité.*

*Une optimisation stochastique pour la correspondance stéréo dense

1. Introduction

Dense stereo matching allows extraction of depth physical properties from two images of the same scene. Images are noted $I_1(x,y)$ and $I_r(x,y)$, where $x \in \{1,\ldots,n_x\}$ and $y \in \{1,\ldots,n_y\}$ denote columns and lines respectively, and I_l and I_r being the left-hand and right-hand side images. Two cameras which observe the same scene have a common field of vision. A point in the scene is projected on the two images in two points which have different coordinates due to the different points of view of the cameras. We express this shifting by horizontal and vertical disparities in the two directions (left-right and right-left): $dx^a(x,y) = x_b - x_a$ and $dy^a(x,y) = y_b - y_a$; $(a,b) \in \{(l,r),(r,l)\}$.

The two cameras do not observe exactly the same physical points in the scene, because some points can be out of the common field of vision or hidden by objects. A point observed by a single camera is declared as an occlusion in the image where it appears. Occlusions are noted $O_l(x,y)$ and $O_r(x,y) \in \{0,1\}$, where 0 stands for an occlusion.

The points that are seen by both cameras can be matched. Disparities can be calculated by this point matching. Having the disparities, depth can easily be computed by triangulation using the cameras and the stereo rig parameters.

Many works have been made in dense stereo matching such as correlation (Sunyoto, Van der Mark, Gavrila, 2004; Faugeras, Keriven, 1998; Jia et al. 2004), and mean field annealing (Huang, Liu, 1997; Huq, Abidi, B, Abidi, M, 2007). The aim of this work is to obtain better results than correlation by adding a global aspect to the stereo vision problem, the correlation inherently having a local aspect, and make this module cooperate with other early vision modules possible, for instance shape from shading and photometric stereo. In order to attain our objective, we formulate the problem in a stochastic optimization approach. The stereo matching is an ill-posed problem (Bertero, Poggio, Torre, 1988) which solution is hidden by local minima and where the notion of gradient does not exist. The most suitable algorithm is simulated annealing, S.A. algorithm, (Aarts, Laarhoven, 1987). For this purpose, stereo matching constraints need to be modeled by energy functional and elementary transformations.

2. Simulated annealing

Simulated annealing is a resolution method for general non linear optimization problems. It is suitable for large size problems, where the global minimum is hidden by several local minima and where the notion of gradient has no signification. This algorithm is applied to a combinatorial discrete problem. Metropolis (Metropolis, Rosenbluth, Teller, 1953) first incorporated these kinds of principles into numerical computations. In order to use the S.A. algorithm, 6 conditions are required (Press and al. 2003) :

1. A description of all possible system configurations $x = \{x_i\}$
2. An initialization $x^0 = \{x_i^0\}$
3. A generator of random changes $x \to x'$
4. An energy functional $E(x)$ that describes the problem and which minimization is the goal of the procedure
5. A high starting temperature T_0
6. A temperature decrease rule $T \to T'$

2.1. Algorithm

1) Begin with initial configuration $x = x_0$ at a high start-ing temperature $T = T_0$.
2) Research the thermal equilibrium at the temperature T.
   a) Propose a random change of the configuration $x \to x'$
   b) Evaluate the energy differential due to this change: $\Delta E = E(x') - E(x)$
   c) i) If the energy decreases $\Delta E \le 0$ then accept the change $x = x'$. ii) If the energy increases $\Delta E > 0$ then accept the change $x \to x'$ with the probability $e^{-\Delta E/T}$.
   d) Repeat a) to c) until the thermal equilibrium is reached at the current floor temperature.
3) Decrease the temperature $T = T'$.
4) Repeat 2) to 3) until the system reaches the minimum.

In practice, we compare a random number between 0 and 1 with the Boltzmann exponential $e^{-\Delta E/T}$, and we accept the change if this number is smaller than this probability. The system reaches the thermal equilibrium after a certain number M of elementary transformations. The number M is set to 100 by some authors. This number should rather depend on the size of the definition domain of the variables. If we try 100 transformations on a variable, there is a difference whether this variable can take one value out of 100, or one value out of 10000 values. In the first case, each value is tried once in average, whereas, in the second case, only 1% of the values is used.

We propose to define M according to the size of the definition domain: $\tau(SizeofDomain)$. The parameter τ is determined by experimentation. The system is considered at high temperature if more than 10% of M attempts are accepted, otherwise, it is at low temperature. If the system is at high temperature, it is considered in thermal equilibrium after 10% of M successful transformations. We get to the next temperature floor by multiplying the control parameter T with a coefficient α : $T = \alpha T$; $0 << \alpha < 1$. For a temperature decrease rule, we take $\alpha = 0.93$ at high temperature and $\alpha = 0.96$ at low temperature. The procedure stops when no more transformations are accepted.

3. Stereo matching

We are only interested in stereo matching general constraints: the resemblance of the primitives, the epipolar lines, the continuity of the disparities and the uniqueness. These constraints are expressed as energy functional and elementary transformations. We consider the correspondences in both directions (left to right and right to left), in order to apply the continuity criterion to the two images and the validity criterion given by the uniqueness constraint for the matching. We handle the disparities and occlusions of the two images in a consistent way.

3.1. Resemblance constraint

This constraint allows the matching of two points only if they have the same grey level. In the case of lambertian surfaces, a point in the scene has the same grey level in both images. A point is a corner of a pixel for which we attribute the average of the grey levels of its four adjacent neighbors:

$$A(x, y) = \sum_{k=0}^{1}\sum_{l=0}^{1} I(x + k, y + l)$$

We express this constraint by the square of the difference of the grey values associated with the matching candidates:

$$\left(A^l(x, y) - A^r\left(x + dx^l(x, y), y + dy^l(x, y)\right)\right)^2$$

This term is weighted with the term representing occlusions, because if there is any occlusion there will be no matching. The energy associated with this constraint over the whole image is:

$$E_R = \sum\sum\left(A^l(x, y) - A^r\left(x + dx^l(x, y), y + dy^l(x, y)\right)\right)^2 O^l(x, y)$$

3.2. Epipolar constraint

A point in the scene is projected on a point in the image. All the points in the scene which belong to the line going through the point of the scene and the optical center are projected on the same point in the image. The projection of this line on the other image represents the epipolar line associated with the point. A point can only be matched with a point in the other which lies on its epipolar line. The search space is reduced to the epipolar line. It is the only constraint that is due to stereoscopic geometry. It is expressed in the elementary transformations.

3.3. Continuity constraint

We consider the physical surfaces as locally continuous. In this case, their euclidean projection is continuous as well. The disparity varies a little, except on the edges; that is the reason why we cancel the continuity constraint at the edges, as well as we cancel the constraint on occlusions, where no matching exists. This constraint is expressed by the energy functional given hereafter. Horizontal and vertical edges are respectively denoted by c_x and c_y: $c_x^a(x,y), c_y^a(x,y) \in \{0,1\}$, where 0 stands for the presence of an edge.

$$E_C^a =$$

$$\sum_{x=1}^{n_x^a} \sum_{y=1}^{n_y^a} \left(\Delta dx_h^a(x,y)\right)^2 c_x^a(x,y) O^a(x,y) O^a(x+1,y) +$$

$$\sum_{x=1}^{n_x^a} \sum_{y=1}^{n_y^a} \left(\Delta dx_v^a(x,y)\right)^2 c_y^a(x,y) O^a(x,y) O^a(x,y+1) +$$

$$\sum_{x=1}^{n_x^a} \sum_{y=1}^{n_y^a} \left(\Delta dy_h^a(x,y)\right)^2 c_x^a(x,y) O^a(x,y) O^a(x+1,y) +$$

$$\sum_{x=1}^{n_x^a} \sum_{y=1}^{n_y^a} \left(\Delta dy_v^a(x,y)\right)^2 c_y^a(x,y) O^a(x,y) O^a(x,y+1)$$

3.4. Occlusions counterweight

With occlusions all over the image, the energy is minimal (null), but this is not a viable solution. This is why we add a counterweight in the energy functional in order to limit the number of occlusions. For one image, this energy functional is:

$$E_O^a = \sum_x \sum_y \left(1 - O^a(x,y)\right)$$

3.5. Uniqueness constraint

Every point can only be matched with at most one point from the other image. This constraint was introduced by Marr (Marr, Poggio, 1979), and it is strictly verified in the case of the non transparent objects.

This constraint simplifies the computations and enables the application of the validity criterion for a given matching. A matching is valid if both disparities (left to right and right to left) enable us to reach the point in the right hand side image from the point in the left hand side image and go back to the left hand side image initial point. This constraint is expressed in the elementary transformations.

3.6. Energy functional

The energy functional of the stereo matching problem used by the S.A. algorithm is the weighted sum of the energy functionals derived from the stereo matching constraints:

$$E = \rho_R E_R + \rho_C (E_C^l + E_C^r) + \rho_O (E_O^l + E_O^r)$$

where ρ terms denote the weights. It is always possible to set one of them to 1 and determine the others accordingly.

3.7. Elementary transformations

The random changes generator is based on the definition of a set of elementary transformations with a decision rule for their application. Figure 1 illustrates the elementary transformations, with the states before and after their application. What is shown is two conjugate epipolar lines. Any complex transformation could be obtained by combining these elementary transformations

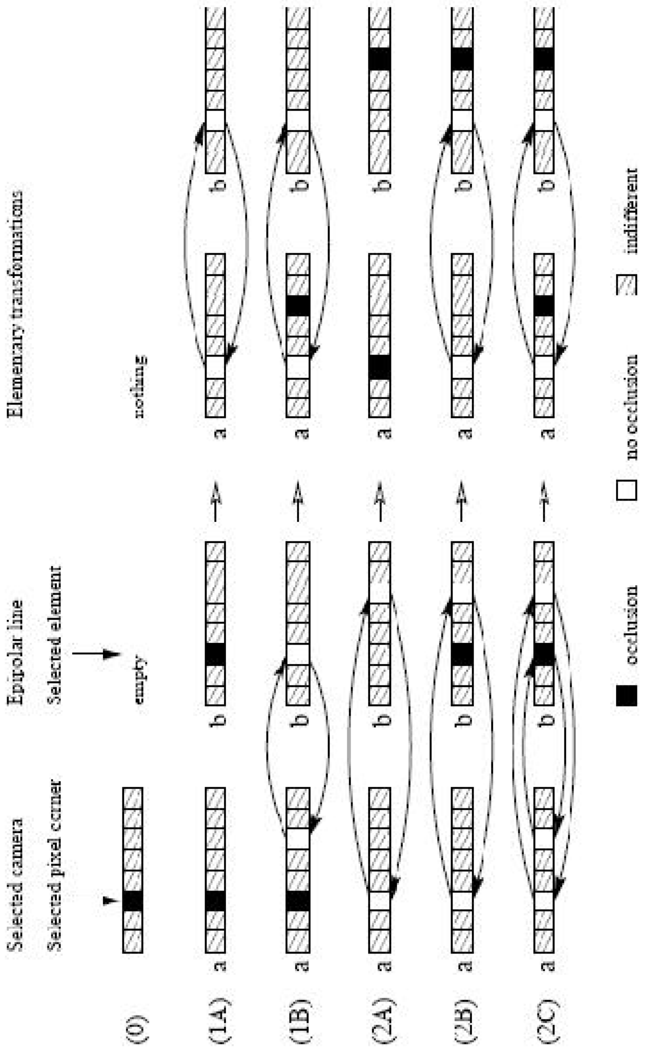

Figure 1 : *Elementary transformations.*

a. Decision rule

We select with equal random one of the two images and a first point. If this point has an occlusion, one of the transformations (0), (1A) or (1B) is applied. If its epipolar line is empty, (0) is applied, otherwise, a point in the epipolar line is selected with equal random for a potential matching. If the selected point has an occlusion, (1A) is applied, otherwise, we apply (1B). If the first point does not have an occlusion, we select with equal random either (2A) or {(2B),(2C)}. If the choice is {(2B),(2C)}, a point in the epipolar line is selected with equal random. If the selected point has an occlusion, (2B) is applied, otherwise (2C) is applied.

b. Transformations

- (0): do nothing.
- (1A): matches the first point with the selected point in its epipolar line.
- (1B): a new matching between the first point and the selected point in its epipolar line; the former correspondent is occluded.
- (2A): establishes two occlusions, one for the first point, the other for its correspondent.
- (2B): occludes the first point former correspondent and establishes a new matching with the selected point in the epipolar line.
- (2C): establishes two occlusions, one for the first point former correspondent and the other for correspondent of the selected point in the epipolar line; a new matching between the first point and the selected epipolar line point is set.

The energy differential is computed locally. The configuration space corresponds to the configurations where the epipolar line and the uniqueness constraints as well as the validity criterion are respected.

4. Correlation

To make a fair comparison between the two approaches, correlation and S.A. algorithm, we have used the same resemblance constraint, epipolar constraint, uniqueness constraint, and validity criterion. The correlation consists in sweeping the candidate point over the epipolar line and to establish a matching with the point that minimizes the resemblance constraint. This is done in both directions (left to right and right to left). Only valid matches are kept. Invalid matches are cancelled and occlusions are set instead.

5. Experimental results

A simple method is used to compute the weighting coefficients in the energy functional. If a tolerance of 0.5 is allowed on the resemblance constraint, the matching is cancelled if the grey level difference is higher than this value: $0.5 \times \rho_R \geq 2 \times \rho_O$. With $\rho_O = 1$, we have $\rho_R \geq 4$.

Furthermore, we assume, for a smooth surface that the difference in disparities is not higher than 1. If the accumulated disparities differences in 4 directions for a pair of matched points are larger than the energy provided by 2 occlusions, the 2 occlusions are set: $2 \times 4 \times \rho_C \geq 2 \times \rho_O$. With $\rho_O = 1$, we have $\rho_C \geq 0.25$.

Finally, as we handle the two cases (left to right and right to left) altogether, we should set: $\rho_O = 2$.

The results obtained with the weighting gains given above are already better than the ones obtained with the correlation. Even better results could be obtained with thorough experimentations on the weighting gains. The starting temperature T_0 is set to 2, which corresponds to the system's state where, virtually, all the elementary transformations are accepted. The parameter τ is set to 100. This parameter's value has proven to give good results.

We experimented our algorithm using both synthesized and real images. Figure 2 shows results on synthesized images.

The figure shows the grey level stereo image pair. The images are produced by an image synthesis system developed for the purpose of verifying and analyzing an image analysis system. We considered 60x40 images to show the point aspect of the image. The data we were interested in are the depth physical properties.

The depth maps generated by the image synthesis system (second row) and therefore corresponding to the solution are used to show the matching errors produced by the methods we are considering. The third and fifth rows show depth maps produced by the correlation and the S.A. algorithms respectively: black pixels stand for occlusions while grey shades correspond to different depths. Dark areas are closer while lighter ones are farther. The fourth and sixth rows show the error maps (difference between the produced depth map and the actual depth map generated by the image synthesis system): black pixels stand for areas where no measurements could have been done; dark areas represent large errors while lighter ones represent small or no errors.

The correlation produced a large number of occlusions, due to a large number of invalid matches. The depth map contains too many errors. The correlation disparity map has been used as an initialization for the S.A. algorithm. This really does not influence the S.A. algorithm since the starting temperature T_0 is high enough to allow all the transformations to be accepted at the beginning, thus moving the system far way from the initialization. Another way to initialize the system could have been by setting up occlusions everywhere and initializing all the disparities to 0. The results generated by the S.A. algorithm show a small number of occlusions. The areas where depth is not correctly produced are less important than for the correlation.

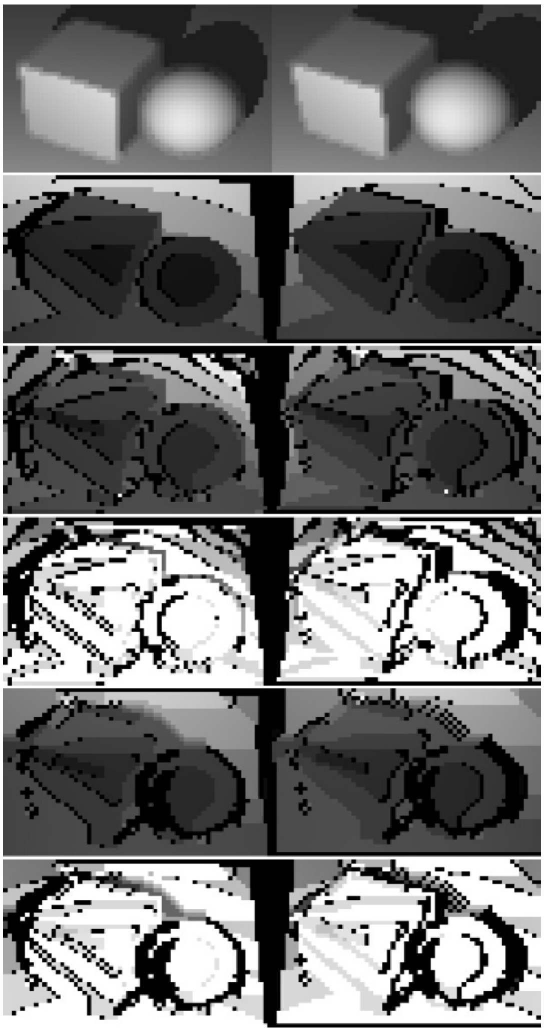

Figure 2: *Experimental results on synthesized imagery.*

Experimental results on real imagery are shown in figure 3. In this example, 4 lighting conditions were used to generate the depth maps. First row shows one of four stereo image pairs. Second row shows the depth maps computed by correlation. Third row shows the depth maps computed using our energy model and the S.A. algorithm.

6. Conclusion

We have tackled the dense stereo matching problem in an optimizational approach. The S.A. algorithm produces better results than the equivalent correlation algorithm. The stereo vision module that we have developped is meant to cooperate with the stereo photometry module (shape from shading), where the shape of the objects is used as a con-straint and the grey level for the resemblance constraint are replaced by the photometric characteristics of the Phong model (Phong, 1975), taking into consideration the grey level changes as a function of the position of the cameras.

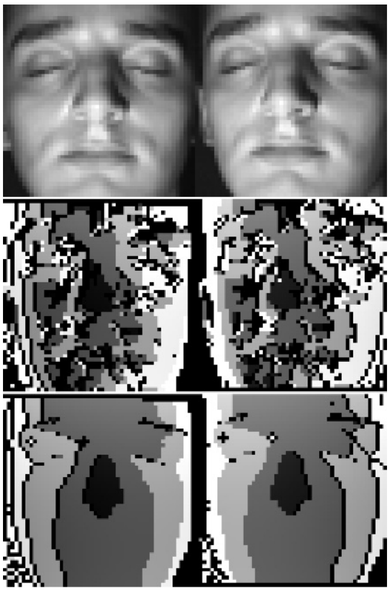

Figure 3: *Experimental results on real imagery.*

7. References

Aarts, P.J.M., Laarhoven, V. 1987. *Simulated Annealing and Applications; Mathematics and its Applications.* D. Reidel Publishing Company

Benameur, W. 2004. Computing the Initial Temperature of Simulated Annealing; Computational Optimization and Applications, v.29 n.3, p.369-385, December 2004.

Bertero, M., Poggio. T. 1988. V. Torre; Ill-Posed Problems in Early Vision; Proceedings of «the IEEE», p.869-889

Faugeras, O., Keriven, R. 1998. Complete Dense Stereo Vi-sion Using Level Set Methods. *Lecture Notes in Com-puter Science*, p.379-398

Huang, J.S., Liu, H.C. 1997. Stereo Vision Using a Microcanonical Mean-Field Annealing Neural-Network. *Neural Computations*, v. 8, No. 1, p.87-104

Huq, S., Abidi, B., Abidi, M. 2007. Stereo-based 3D Face Modeling using Annealing in Local Energy Minimiza-tion; ICIAP, p.265-272

Jia, Y. and al. 2004. *A Miniature Stereo Vision Machine (MSVM-III) for Dense Disparity Mapping.* ICPR, p.728-731

Marr, D., Poggio, T. 1979. A Computational Theory of Human Stereo Vision. Proceedings of the Royal Society, B.204, p.301-328

Metropolis, N., Rosenbluth, K., Teller, **A. 1953.** *Journal of Chemical Physics*, v. 21, p.1087-1092

Phong, B. 1975. Illumination for Computer Generated Pictures. Communications of «the ACM», v.18, n.6, p.311-317

Sunyoto, H., van der Mark, W., Gavrila, **D.M. 2004.** A Comparative Study of Fast Dense Stereo Vision Algorithms. «Intelligent Vehicles Symposium», IEEE, Volume 14, Issue 17, p.319-324

W. Press and al. 2003. *Numerical Recipes in C: the Art of Scientific Programming*. Seconde Edition. Cambridge University Press