AREA MATCHING BASED ON BELIEF PROPAGATION WITH APPLICATIONS TO FACE MODELING

Davide Onofrio, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica e Informazione - Politecnico di Milano
Piazza Leonardo Da Vinci 32, 20133 Milano, Italy

ABSTRACT

In the past few years several systems for face recognition and identification based on the analysis of 3D data have been proposed. In order for such systems to be of practical use, the 3D data extraction process is expected to be fast and reliable. In this paper we propose a general approach to multi-view 3D surface reconstruction, specialized for the specific problem of face modeling. Our method determines correspondences between surface patches on different views through a modeling of disparity maps based on Markov Random Fields (MRFs). In order to reduce the occurrence of outliers the MRF-based modeling is bound to satisfy the epipolar constraint. We apply the belief propagation algorithm to the MRF model in order to perform a maximum-a-posteriori estimation of such correspondences.

1. INTRODUCTION

Systems for face recognition are steadily evolving towards the analysis of 3D data, as it seems to offer more reliable classification solutions. The performance of such systems is strongly dependent on a fast and reliable acquisition of 3D data relative to human faces. When dealing with solutions for multi-view 3D modeling, one crucial step is that of correspondence matching. This problem is generally approached using solutions that are closely related to optical flow estimation [1]. In fact, such methods use the brightness constancy constraint to determine correspondences between points in multiple views. This assumption, indeed, is a strong one as it implies that the viewed surfaces be Lambertian and that the illumination be diffuse. Nonetheless, the brightness constraint is widely known to provide reliable results for the considered problem.

Like any optical flow method, we need further constraints in addition to the brightness constancy. The simplest choice would be to consider a smoothness constraint on the correspondence field. This solution, however, does not consider several problems that we need to face in depth estimation, and this is particularly true for face modeling. Face modeling, in fact, suffers from problems of occlusions and lack of image structure. The former problem prevents us from finding correspondences, while the latter makes the brightness constraint useless.

Our MRF-based approach offers a promising way to overcome these difficulties [2] as it provides us with an additional smoothness constraint while dealing with occlusions and texturing problems.

As we know, the brightness constraint applied to textureless regions gives us no information. For this reason, we adopt a MRF model to propagate information contained in highly textured neighboring regions. We do so by modeling the disparity field using a MRF whose Gibb's energy is made of two terms: one accounts for brightness constancy, while the other (clique potential) describes the diffusion of information from neighboring regions. As a direct computation of the marginal probabilities in the MRF model is generally a computationally intensive task (the complexity grows exponentially with the number of grid nodes), we adopted a fast algorithm to compute them, based on belief propagation (BP).

As far as occlusions are concerned we can adopt two different types of countermeasures. One is inherent in the modeling process, as it consists of modifying the shape of the belief function in order to account for depth discontinuities. The other solution is associated to the acquisition geometry and is specific of the application scenario that we are considering (face modeling). In fact, we use a calibrated trinocular system whose epipolar geometry is known and can provide us with additional constraints to reduce outliers. In addition, we work on feature correspondence estimation in a pairwise fashion,

This work was developed within the FIRB-VICOM project (www.vicom-project.it) funded by the Italian Ministry of University and Scientific Research (MIUR); and within the VISNET project, a European Network of Excellence (www.visnet-noe.org)

0-7803-8554-3/04/$20.00 ©2004 IEEE.

because many of the face details are only visible to two of the three cameras, therefore we can devise a suitable 3D fusion process to integrate information coming from such pairs.

2. TRINOCULAR 3D RECONSTRUCTION

What stated until now has a general validity, therefore applies also to the trinocular calibrated camera system that we used in order to acquire face images. In our system one of the cameras, taken as a reference (*master*) has a reasonable frontal, occlusions free, view, while the others (*slaves*) show some occlusions. First the three cameras are calibrated [3], then the images acquired synchronously can be processed. The proposed algorithm computes two dense disparity maps betweeen the *master* and the other two *slaves* views each map is modeled by one pairwise MRF [4]. We estimate marginal probability by performing belief propagation iterations on the MRF.
At the end of the process, we coupled the two MRF results in order to satisfy epipolar constraint on the triplet of images and hence to eliminate outliers.

2.1. Preprocessing

The acquired triplets are low-pass filtered to reduce the effect of noise. For the *master* view, the Euclidean norm of gradient is calculated:

$$|\nabla I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \tag{1}$$

in order to establish patch size for next steps. Starting from a minimum patch size, in the *master* view, we expand the patch until $|\nabla I|$ reaches a prefixed threshold or the size reaches its maximum value. This shape adaptive window provides high reliability around strong feature where high values of $|\nabla I|$ are expected.

2.2. Stereo matching with MRF's

2.2.1 MRF description
A node *i* on the MRF grid represents the disparity d_i of the pixel in position w_i on the *master* image, neighbour nodes represent neighbour pixels, we propose to calculate disparity fields by minimizing an energy expression involving a fidelity term and a smoothing term:

$$E(\{d\}) = \sum_i \iint_{W+[w_i]} (I_{master}(w) - I_n(w+d_i))^2 d^2 w + \lambda \cdot \sum_{(ij)} \psi_{ij}(d_i, d_j) \tag{2}$$

where the sum on *i* is taken on the nodes of the grid and the sum on *ij* is over nearest neighbors on the lattice, W is a bounded set around the point w_i, I_{master} and I_n are respectively the brightness of the master view and of one of the *slaves* images with respect to the disparity is calculated, λ is a regularization parameter which balances the influence between the two terms of Eq. (2).
The choice of the function ψ_{ij} is made in order to encourage smoothing within a region. Among the many diffusivity function proposed [5] we adopted that suggested by Geman and McClure [6]:

$$\psi(t) = \frac{t^2}{1+t^2} \tag{3}$$

Using this expression for ψ_{ij} in Eq. (2) we have:

$$\psi_{ij}(d_i, d_j) = \frac{(d_i - d_j)^2}{1+(d_i - d_j)^2} \tag{4}$$

Assuming:

$$\chi_i(d_i) = \iint_{W+[w_i]} (I_{master}(w) - I_n(w+d_i))^2 d^2 w \tag{5}$$

for every node *i*, we permit d_i to assume values only on the epipolar line on the *slave* image of the point w_i on the *master* image, χ_i therefore is a function of one variable. The size of the patch W on which integration is performed changes from node to node as described in previous section.
If we define 'evidence' as

$$\phi_i(d_i) = e^{-\chi_i(d_i)/T} \tag{6}$$

and define the "compatibility" between the disparities at neighboring nodes by

$$\gamma_{ij}(d_i, d_j) = e^{-\lambda \cdot \psi_{ij}(d_i, d_j)/T} \tag{7}$$

where T is assumed to be constant equal to 1. The probabilitiy distribution of a disparity map d_i will factorize into a product function of the maximal cliques of the graph.

$$p(\{d\}) = \frac{1}{Z} \prod_{(ij)} \gamma_{ij}(d_i, d_j) \prod_i \phi_i(d_i) \tag{8}$$

Z is a normalization constant. Thus in the Markov model, ψ_{ij} and χ_i are the potential functions of the Gibbs energy and minimizing $E(\{d\})$ as defined in Eq. (2) corresponds for the Eq. (8) to statistical MAP estimation [7]:

$$p(\{d\}) = \frac{1}{Z} e^{-E(\{d\})/T} \tag{9}$$

2.2.2 Belief propagation
In the BP algorithm [8] we introduce variables such as $m_{ij}(d_j)$ which can be understood as a "message" from an hidden node *i* to the hidden node *j* about what state node *j*

should be in, see Fig. 1. The message $m_{ij}(d_j)$ will be a vector of the same dimensionality as d_j, with each component being proportional to how likely node i thinks it is that node j will be in the corresponding state.
To start BP we need to calculate the 'evidence' and the 'compatibility' functions to update iteratively the messages. To perform BP first, we initialize all messages $m_{ij}(d_i)$ as uniform distribution; second we update $m_{ij}(d_i)$, iteratively for $k=1:K$,

$$m_{ij}^{(k+1)}(d_j) = \sum_{x_i} \phi_i(d_i)\psi_{ij}(d_i, d_j) \prod_{k \in N(i)\setminus j} m_{ki}^{(k)}(d_i) \quad (10)$$

Note that in the right-hand side, we take the product over all messages going into node i except for the one coming from node j.

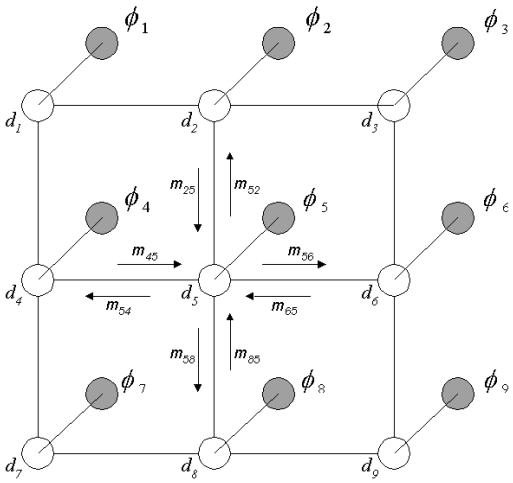

Fig. 1. An illustration of the message passed in BP in a square lattice pairwise MRF

In
the BP algorithm, the belief at a node i is proportional to the product of the local evidence at that node ($\phi_i(d_i)$), and all the messages coming into node i:

$$b_i(d_i) = k \cdot \phi_i(d_i) \prod_{j \in N(i)} m_{ji}(d_i) \quad (11)$$

where k is a normalization constant (the beliefs must sum to 1) and $N(i)$ denotes neighboring i.
Then we compute MAP estimation of disparity d_i as given by BP with:

$$d_i^{MAP} = \arg\max_{d_k} b_i(d_k) \quad (12)$$

2.3. Outlier elimination

Results of the two disparity fields are coupled to discard false correspondences matching. The disparity maps built, automatically, satisfy epipolar constraint between the *master* image and the *slave* image where disparity is calculated to; nevertheless the matched points, on the *slave* images are not guaranteed to stay on the epipolar line of each other. We further check that all the matched points, as outgoing from BP processes, satisfy this last epipolar contraint. If this is not the case we are in presence of outliers and points are rejected.
The process just described also could come automatically with BP: if we add a term in the energy expression, Eq. (2), that grows with the distance between the matched point and the epipolar line on the other *slave* image.

2.4. Back projection

Points that remain after the outlier elimination are back-projected in 3d space to obtain the estimation of the 3D structure of the face.

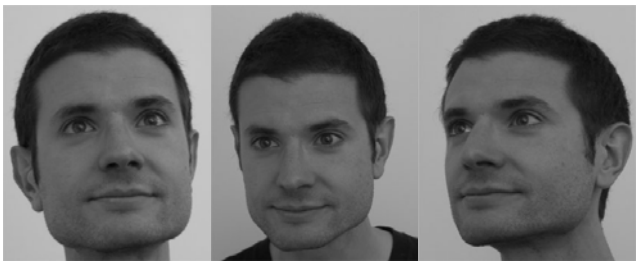

Fig. 2. "Face" image triplet

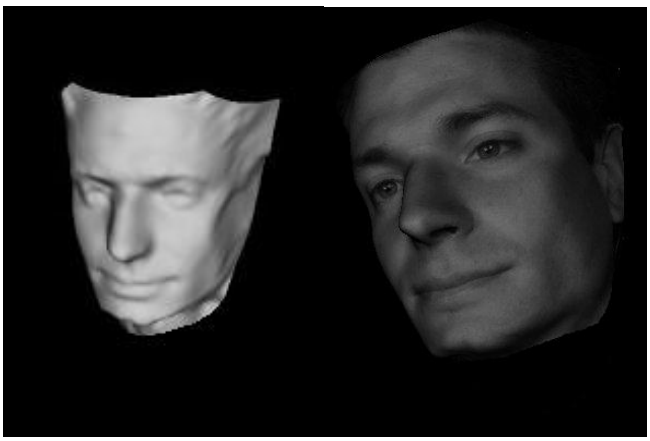

Fig. 3. 3D reconstruction of "Face"

3. EXPERIMENTAL RESULTS

The proposed algorithm is applied to triplets of images acquired with a trinocular calibrated camera system. As an example of triplet of images acquired see Fig. 2. The size of images are 800×600. In Fig. 3 an example of

reconstruction is shown. The parameter λ as defined in Eq. (2) is determined heuristically, optimal values depend on the images we are processing. The parameter can be varied to gain some insight about the scene: for little values of λ in Eq. (2) the '*evidence*' dominates the '*compatibility*' and the resultant map disparity will depend mostly on the first term of Eq. (2) little information about disparity is exchanged between MRF nodes, for high values of λ the *compatibility* term dominates and the disparity map will be smooth. Those matched points that in the two cases change will be due to occlusion, or to textureless regions.

In cyclic graph, like the pairwise MRF's described above, if we ignore the existence of loops and permit the nodes to continue communicating with each other, messages may circulate indefinitely around these loops and the process may not converge to a stable equilibrium [9], we observed experimentally this behaviour, for this we perform a time average operation after a fixed number of iteration. To evaluate the number of iterations needed in the BP algorithm, we present in Table 1 the average number of false matching (where false is intended at distance major than δ_d) as proposed in [10] versus the number of iterations performed: BP algorithm does not need many iterations in the sense of Marr [11].

Table 1. List of bad matching versus iterations.

Iteration	$\frac{1}{N}\sum_i(\lvert\hat{d}_i - d_i\rvert \geq \delta_d)$
1	25.12
8	8.75
16	4.56
32	1.98

4. CONCLUSIONS

Face modeling is a critical task for 3D face recognition systems. In order to perform this task we acquire images by trinocular calibrated cameras and find correspondences between the three views. Moreover, face modeling suffers strongly of occlusions and textureless regions, to deal with this problems we model disparity maps with MRF's, in order to propagate information from textured to textureless regions. We applied the BP algorithm, to obtain the maximum a posteriori estimation of the disparity maps. In order to reduce false matching due to occlusions we eliminate outliers by epipolar constraint check.

As a result the proposed algorithm provides very accurate 3D face model.

One drawback of this implementation is that the parameter λ is found heuristically, one way to overcome this aspect could be that of let λ be a function of the node depending inversely on the modulus of the gradient of the brightness function in the *master* image, in so doing we permit to propagate more information about disparity in homogeneous regions and less around strong features.

One of the virtue of the BP algorithm is that we can compute marginal probabilities, at least approximatively, in a time that grows only linearly with the number of nodes in the system. The BP is an highly parallelizable algorithm, as a future work, we want to speed up execution time going in this direction, further we want to investigate if Generalized Belief Propagation [12] can improve correspondences matching reliability.

6. REFERENCES

[1] F. Pedersini, P. Pigazzini, A. Sarti, S. Tubaro, "3D Area Matching with Arbitrary Multiview Geometry," *EURASIP Signal Processing: Image Communication - Special Issue on 3D Video Technology*, Elsevier, vol. 14, N. 1-2, pp.71-94, October 1998.

[2] S. and D. Geman, "Stocasthic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence* , vol. 6, N. 6, pp 721-741, 1984.

[3] F. Pedersini, A. Sarti, S. Tubaro, "Multicamera Systems: Calibration and Applications," *IEEE Signal Processing Magazine, Special Issue on Stereo and 3D Imaging*, vol. 16, N. 3, pp. 55-65, May 1999.

[4] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.

[5] S. Teboul, L. Blanc-Feraud, G. Aubert and M. Barlaud, "Variational Approach for Edge-Preserving Regularization Using Coupled PDE," *IEEE Trans. Image Processing*, vol. 7, N. 3, pp. 387-397, March 1998.

[6] S. Geman and D.E. McClure, "Bayesian Image Analysis: An Application to Single Photon Emission Tomography," *Proc. Stat. Comput. Sect.*, Washington, DC, pp. 12-18, 1985.

[7] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. On Information Theory*, vol. 18, pp 232-240, 1972.

[8] J. Yedida, W. Freeman, Y. Weiss, "Bethe Free Energy, Kikuchi Approximations and Belief Propagation Algorithms," Technical Report TR-2001-16, *Mitsubishi Electric Research Labs*, 2001.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo, Calif, 1988.

[10] D. Scharstein, R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision*, vol. 47, N. 1, pp 7-42, 2002.

[11] D. Marr, *Vision*, H. Freeman and Co., 1982.

[12] J. S. Yedida, W.T. Freeman, Y. Weiss "Generalized belief propagation. Technical Report 2000-26," *MERL, Mitsubishi Electric Research Labs*, www.merl.com, 2000.