FAST STEREO MATCHING METHOD BASED ON OPTIMIZED CORRELATION ALGORITHM FOR FACE DEPTH ESTIMATION

Amel Aissaoui, Rémi Auguste, Tarek Yahiaoui, Jean Martinet and Chaabane Djeraba
LIFL UMR Lille1-CNRS n 8022, IRCICA, 50 avenue Halley, 59658 Villeneuve d'Ascq, France

Keywords: Face Depth Estimation, Correlation-based Method, Stereovision.

Abstract: In this paper, we introduce a novel approach for face stereo reconstruction based on stereo vision. The approach is based on real time generation of facial disparity map, requiring neither expensive devices nor generic face model. An algorithm based on incorporating topological information of the face in the disparity estimation process is proposed to enhance the result of the 3D reconstruction. Some experimental results are presented to demonstrate the reconstruction accuracy of the proposed method.

1 INTRODUCTION

Face depth estimation is an important problem that was conjointly studied with face animation, facial analysis and face recognition. In the past few decades, many approaches have been proposed, including 3D from stereo matching (Furukawa and Ponce, 2010), 3D morphable model based methods and (Choi et al., 2010), structure from motion (Chowdhury and Chellappa, 2003) and shape from shading techniques (Chow and Yuen, 2009). However, how to efficiently acquire facial depth information from stereo images is still a challenging problem especially for real time application.

So far, several attempts have been made to deal with 3D face reconstruction from stereo images. (Lengagne et al., 2000) proposed a user interactive approach to deform an animated 3D mesh model from two stereo images. They use a priori knowledge and differential constraints on the 3D model to recover the surfaces of facial areas that are not reliably obtained from stereo alone. (Mallick and Trivedi, 2003) use parallel stereo images and a set of manually selected corresponding feature points to compute the rotation and translation matrix that are used to fit the 3D mesh model to the computed 3D feature points. (Cryer et al., 1995) proposed to merge the dense depth maps obtained separately from stereo and Shape From Shading (SFS) in the frequency domain. The merging process is based on the assumption that shape from stereo is good at recovering high frequency information and shape from shading is good at recovering low frequency information. Recently, many methods use improved SFS techniques to enhance the stereo results (Chow and Yuen, 2009). (Zheng et al., 2007), used a reference 3D face as an intermediate for correspondence estimation. The virtual face images with known correspondences are first synthesized from the reference face. The known correspondences are then extended to the incoming stereo face images, using face alignment and warping. In (Wu et al., 2008), authors do not use any external model, the feature correspondences between images are extracted and the disparity map is initialized. Then an iterative algorithm was used to refine automatically the disparity map using other images taken with different baseline.

In this paper, we propose an improved framework for determining the disparity information of a human face from stereo matching in a binocular vision system using correlation based methods. While many stereo matching algorithms have been proposed in recent years (Scharstein and Szeliski, 2002), correlation-based algorithms still have an edge due to speed and less memory requirements (Heo et al., 2011). For the same reasons, we choose to use a correlation based method improved by incorporating the topological information specific to the face obtained by fitting an Active Shape Model (ASM) (Milborrow and Nicolls, 2008) on both images in the initialization step of the algorithm, while maintaining its real-time suitability. Our method demonstrated a satisfactory performance in terms of processing time and point matching accuracy.

The remainder of this paper is organized as follows. First, we describe the principle of depth estimation in Section 2. In Section 3, we present the

Aissaoui A., Auguste R., Yahiaoui T., Martinet J. and Djeraba C..
FAST STEREO MATCHING METHOD BASED ON OPTIMIZED CORRELATION ALGORITHM FOR FACE DEPTH ESTIMATION.
DOI: 10.5220/0003845703770380
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 377-380
ISBN: 978-989-8565-04-4

proposed method. The result of our implementation is given in Section 4. Finally, Section 5 concludes the paper.

2 DEPTH ESTIMATION PROCESS

Depth estimation process in binocular stereo system consists of reconstruction of 3D information of a scene captured from two different points of view (see Figure 1).

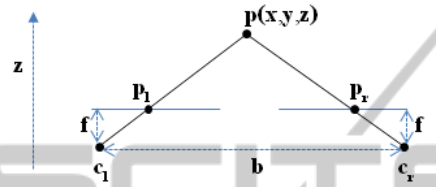

Figure 1: Geometric model of calibrated stereo vision system.

p : is an object point in real word.

p_l **(resp.** p_r**)** : is the correspondence point of p in left (resp. right) image.

f : is the camera focal,

b : is the camera baseline,

As Figure 1 shows and according to similar triangles' principle, the disparity can be formulated as the following generalized form:

$$\frac{z}{f} = \frac{x}{x_l} = \frac{(x-b)}{x_r} = \frac{y}{y_l} = \frac{y}{y_r}. \quad (1)$$

From Equation (2) we have:

$$d = (x-b) = \frac{fb}{z}. \quad (2)$$

Where d is the disparity. Its value increases when the distance between the point p and the camera decreases. To estimate its value, it is necessary to find pixels in both images that correspond to the projection of the same real word point. This process is called *stereo matching*.

A correlation-based stereo matching algorithms typically produces dense depth maps by calculating the matching costs for each pixel at each disparity level in a certain range. Afterwards, the matching costs for all disparity levels can be aggregated within a certain neighborhood window. Finally, the algorithm searches for the lowest cost match for each pixel. The most common is a simple winner takes-all (WTA) minimum or maximum search over all possible disparity levels.

Different cost measures are used in correlation-based methods. The most common ones are: Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), Normalized Cross Correlation (NCC) and Sum of Hamming Distances (SHD).

3 PROPOSED METHOD FOR DISPARITY ESTIMATION

In order to estimate the disparity map, we adopt a correlation based methods because of their low cost in processing time. However, photo-consistency measures used in these methods are not always sufficient to recover precise geometry, particularly in low-textured scene regions, in case of occlusion and large disparity. It can therefore be helpful to impose shape priors that bias the reconstruction to have desired characteristics.

For this purpose, we use an Active Shape Model (ASM) to obtain prior topological information about the face. These information reduce the search area from the entire epipolar line to only a small segment. In other words, given a right point in the nose region of the face, we search only in the same region in the left image. This guarantees the smoothness of the disparity map because a point in a topological region (nose, eye, etc.) in the left image will certainly matched with a point in the same region in the right image. As a consequence, disparities values will be continuous and a pixel in eye region will never exceed another in nose region.

Fitting the ASM on both images determine the coordinates of the main features points in the right and left image which are subsequently used to compute the shift vectors of the corresponding feature points of the face without applying the classical methods discussed in Section (2.1). The shift vectors for non-feature points are then determined using a correlation-based method using the features point's disparities.

3.1 Sparse Disparity Calculation

In order to establish the sparse matching, we start by applying an ASM fitting algorithm on both images. The ASM algorithm aims to match a statistical face shape obtained by an offline training process, to a new face image by updating the model parameters to best match to all features points (see Figure 2). In our method, we used the ASM fitting, not only for detection facial features points, but also as an automatic stereo point matching.

After fitting the ASM, we obtain the 2D coordinates of n face feature points in the right $R =$

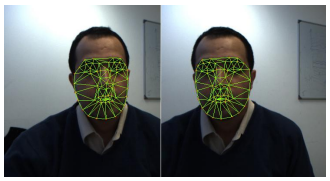

Figure 2: ASM fitting on left and right stereo images.

$(x_i, y_i), i \in [1,n]$ and the left $L = (x_i^{'}, y_i^{'}), i \in [1,n]$ images. Since we use a calibrated system and rectified stereo pairs, the y coordinates of each corresponding points are then normalized to their mean. For each face feature point p_i, the Euclidian distance between its right and left coordinates is calculated to obtain its disparity $d(p_i)$ as follows:

$$d(p_i) = \sqrt{(x_i - x_i^{'})^2 - (y_i - y_i^{'})^2}. \quad (3)$$

After the ASM fitting and the disparity calculation, we now have a set n of features points with 3 coordinates: $P = \{p_i(x,y,d), i \in [1,n]\}$ which will used in the dense disparity calculation step.

3.2 Dense Disparity Calculation

In this step, we calculate disparity of non-characteristic points of the face using the obtained sparse representation.

The first step consists of projecting the face feature points in the 3D space $(o,\vec{x},\vec{y},\vec{d})$ to obtain a 3D ASM for the face. The 3D ASM is then projected on $(o,\vec{x},\vec{d})$ and $(o,\vec{y},\vec{d})$ spaces. This step provides information about the disparity variation on the horizontal and the vertical profiles.

Using the characteristic points projected on both 2D plans $(o,\vec{x},\vec{d})$ and $(o,\vec{y},\vec{d})$, we define the disparity interval of each point $p(x,y)$ according to its neighbor characteristic points that are the left neighbor $p^{LeftNeighbor}$ and the right neighbor $p^{RightNeighbor}$. The disparity interval on $\vec{x}$ axis is defined as $[DispMin_x^p, DispMax_x^p]$, where $DispMin_x^p$ is the disparity of $p^{LeftNeighbor}$ and $DispMax_x^p$ is that of $p^{RightNeighbor}$. In the same way, we define the disparity interval according to y coordinate as $[DispMin_y^p, DispMax_y^p]$. Finally, the disparity interval of $p(x,y)$ is given as shown in the following equation.

$$[DispMin^p, DispMax^p] = [DispMin_x^p, DispMax_x^p] \cup [DispMin_y^p, DispMax_y^p]. \quad (4)$$

In the second step, we calculate the disparity of all non-characteristic points, using their disparity intervals to initialize the algorithm of the correlation, to obtain the dense disparity map.

Given a left image point p_l, a correlation window w and a disparity interval $[DispMin^{p_l}, DispMax^{p_l}]$, we aim at obtaining the disparity $d \in [DispMin^{p_l}, DispMax^{p_l}]$, which maximizes the correlation equation $E(d)$:

$$E(d) = Similarity(p_l(x,y), p_r(x+d,y)) \quad (5)$$

For the similarity function, we have used the SAD measure (Hirschmuller, 2001) that it is calculated by subtracting pixel grey level values within an $n*m$ rectangular neighborhood window w between the reference image I_l and the target image I_r followed by the aggregation of absolute differences within the square window.

$$SAD_{I_l(x,y),I_r(x^{'},y^{'}))} = \sum_{u=0}^{m}\sum_{v=0}^{n} |I_l(x+u,y+v) - I_r(x^{'}+u+d,y^{'}+v)|. \quad (6)$$

4 RESULTS AND DISCUSSION

In this section, we describe our implementation and our results. We use, in our work, a Bumblebee stereoscopic system composed of two CDD pre-calibrated cameras mounted on a horizontal support.

In Figure 7, we compare the disparity map estimated by the SAD method to our method that includes the prior knowledge by applying ASM. Results show that integrating prior knowledge about face can enhance the disparity map in terms of smoothness and also in terms of reducing the missing data named *holes* (or noise) occurring from uncertain disparities.

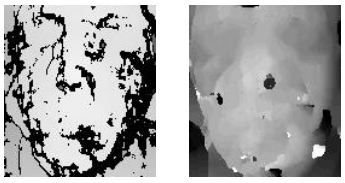

Figure 3: Disparity map: (a) SAD correlation based methods (window 11x11), (b) Our method (SAD+ASM).

In order to obtain the 3D model of the face, The depth map is generated using Equation (2) in Section 2 and preprocessed by applying an ellipsoid mask to crop the face region. In order to fill holes, we applied a selective median filter (with a $7*7$ kernel size), which is often used to preprocess depth data. A point cloud for the face is then generated and the texture

Figure 4: Texture mapping and 3D model generation.

mapping is performed using the OpenGL library of computer graphics.

The results show that the proposed strategy, consisting of incorporating prior knowledge in the disparity estimation process, is robust and accurate. It improves the result of general correlation based methods by considering the face shape and its topological regions, while maintaining its real-time suitability.

5 CONCLUSIONS AND PERSPECTIVES

This paper presents an original attempt to a practical face depth estimation in passive stereoscopic system. Unlike other general methods used for disparity calculation for any object, we introduced a specific method for depth estimation of face that uses the shape characteristics of the human face, obtained by adjusting the form of an active model, to improve result of the correlation-based method. Our method enhanced the classical correlation based method for disparity calculation, in terms of depth estimation efficiency, with maintaining its real-time suitability. The experimental results show that the proposed algorithm produces a smooth and dense 3D point cloud model of human face, applicable to a wide range of real-time 3D face reconstruction situations.

Our approach also opens up many perspectives for improvement and expansion. The estimation of the sparse disparity can be improved by using other versions of the Active Shape Model used in our work. For instance, Active Appearance Models (Cootes et al., 2001) is likely to give more successful adjustments because they use the texture information or the 3D Active Appearance Models (Xiao et al., 2004) which is robust to pose variation. The symmetry propriety of the face can also be incorporated in the estimation process to further improve the results. Finally, it would be interesting to test our method on a stereo face database with ground truth. However, existing databases usually contain scenes and objects. For this, we plan to create a specific database of stereoscopic faces with a ground truth to evaluate our method in a complete way.

REFERENCES

Choi, J., Medioni, G., Lin, Y., Silva, L., Regina, O., Pamplona, M., and Faltemier, T. (2010). 3d face reconstruction using a single or multiple views. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3959 –3962.

Chow, C. and Yuen, S. (2009). Recovering shape by shading and stereo under lambertian shading model. *International journal of computer vision*, 85(1):58–100.

Chowdhury, A. K. R. and Chellappa, R. (2003). Face reconstruction from monocular video using uncertainty analysis and a generic model. *Computer Vision and Image Understanding*, 91:188–213.

Cootes, T., Edwards, G., and Taylor, C. (2001). Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):681–685.

Cryer, J., Tsai, P., and Shah, M. (1995). Integration of shape from shading and stereo* 1. *Pattern recognition*, 28(7):1033–1043.

Furukawa, Y. and Ponce, J. (2010). Accurate, Dense, and Robust Multiview Stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1362–1376.

Heo, Y. S., Lee, K. M., and Lee, S. U. (2011). Robust stereo matching using adaptive normalized cross-correlation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:807–822.

Hirschmuller, H. (2001). Improvements in real-time correlation-based stereo vision. In *Stereo and Multi-Baseline Vision, 2001.(SMBV 2001). Proceedings. IEEE Workshop on*, pages 141–148. IEEE.

Lengagne, R., Fua, P., and Monga, O. (2000). 3d stereo reconstruction of human faces driven by differential constraints. *Image and Vision Computing*, 18(4):337–343.

Mallick, S. P. and Trivedi, M. (2003). Parametric face modeling and affect synthesis. In *Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 2*, ICME '03, pages 225–228, Washington, DC, USA. IEEE Computer Society.

Milborrow, S. and Nicolls, F. (2008). Locating facial features with an extended active shape model. *Computer Vision–ECCV 2008*, pages 504–513.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47:7–42.

Wu, X., Dai, C., and Liu, J. (2008). A novel approach for face recognition based on stereo image processing algorithm. In *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on*, pages 1245 –1249.

Xiao, J., Baker, S., Matthews, I., and Kanade, T. (2004). Real-time combined 2d+3d active appearance models. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 535 – 542.

Zheng, Y., Chang, J., Zheng, Z., and Wang, Z. (2007). 3d face reconstruction from stereo: A model based approach. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, pages III –65 –III –68.