Stereoscopic face images matching

Martin Klaudíny*

Department of Computer Graphics and Multimedia
Brno University of Technology
Brno, Czech Republic

Abstract

This paper is dedicated to the problem of face images matching in passive stereoscopic photogrammetry. The aim of the presented work was to develop a method for correspondence search in a pair of high-resolution images which allows a reconstruction of high-quality 3D face model. The proposed technique combines global approach to the construction of a disparity map based on a graph cut with fast local method. The initial estimate of solution by local approach is used for reduction of a disparity space. Final disparity map is determined by single minimum cut in the reduced graph with 3D grid topology. The reconstructed 3D model of the face has good quality similar to the result by purely global approach. However, the computational time and memory consumption are significantly smaller in the proposed technique comparing to purely global approach.

Keywords: 3D face capture, passive stereoscopic photogrammetry, stereoscopic matching, window-based correspondence, graph cut, maximum flow, disparity range reduction

1 Introduction

This work belongs to the field of 3D capture. *The 3D capture of face* has specific constraints and requirements with respect to a character of examined object. The issues connected with a safety, speed and natural behaviour of a person make the face capture system more complex than general-purpose capture systems. Because of these requirements, the structured light techniques or the techniques based on *a stereoscopic photogrammetry* are usually used. The majority of methods from both groups project some kind of pattern on the captured face. A pattern projection is crucial for a calculation of depth in structured light techniques. Active photogrammetry techniques use the projection of random pattern to improve the quality of matching between the views. Main disadvantage of using patterns is the acquisition of shape without an appearance. This problem has been overcome by projecting pattern in the IR part of electromagnetic spectrum [14]. However, this implies complex capture rig with specialised equipment. Another limitation is strict regulation of scene illumination to avoid contamination of a pattern.

A research challenge is to develop a capture system without active lighting components which has same accuracy and efficiency as the systems using pattern projection. The advantages of such system are simpler capture rig and less restrictive constraints on an illumination in a scene. This step would lead to the transition of 3D face capture systems from laboratory conditions to real environment. The challenge is pursued by passive stereoscopic photogrammetry techniques. The matching is performed on usual images of captured object. Therefore, the algorithms have to be more sophisticated to achieve the quality of result similar to previously mentioned groups of techniques.

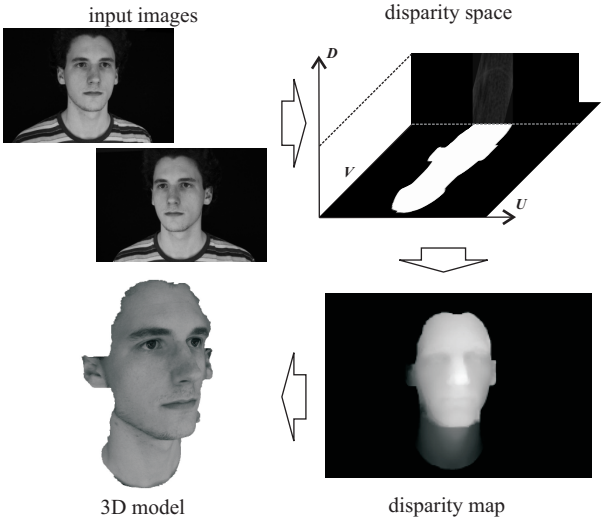


Figure 1: The scheme of the 3D face capture.

This paper addresses a problem of face images matching in the passive stereoscopic photogrammetry. The goal was to develop the correspondence method which allows the reconstruction of high-quality 3D face models. Input images provided by the built capture rig have high resolution (8.2 megapixels). This offered an opportunity to investigate whether such resolution can bring enough skin details to overcome traditional difficulties with matching pattern-free images of the face.

Figure 1 schematically illustrates a processing in the 3D

*xklaud00@stud.fit.vutbr.cz

capture. The processing pipeline has same stages regardless of captured object. However, the algorithms can exploit a priori knowledge that the application is a human face. The input images obtained by calibrated capture rig are pre-processed to simplify following computation. *A correspondence search* finds the pairs of corresponding pixels in the images. It can be divided into two main phases - *a construction of disparity space* and *a construction of disparity map*. The first phase records a matching score for all relevant pairs of pixels in images to a disparity space. The second phase extracts correct matching between pixels from the disparity space in a form of disparity map. The textured 3D model is reconstructed from the disparity map which can also be seen as a range image.

The correspondence search, which is crucial task in the 3D capture, can be solved by large number of methods. They are considered as local or global according to an extent of data used for matching one pixel [10]. The local techniques [14] are fast but inaccurate for the materials with weak texture such as human skin. The global techniques treat the construction of disparity map as an optimisation problem. Therefore, they achieve better quality of disparity map but computational demands are high. The main representatives are simulated annealing, probabilistic diffusion, dynamic programming [11] and graph cuts while the methods based on graph cuts achieve the best results [10]. The part of them use the alpha-expansion algorithm which iteratively applies the graph cut on the 2D grid graph to minimise an energy function [6]. The energy function can define complex smoothness term but the algorithm does not quarantee global minimum. On the other hand, the techniques such as [8, 5] compute optimal solution by single cut on the 3D grid graph which allows only linear smoothness term. To cope with long computational time and huge memory consumption of the mentioned graph cut techniques, several approaches were proposed [9, 13]. They reduce a search space through hierarchical computation of graph cut, initial rough estimate of the solution, or the set of best matches for each pixel.

The technique presented in this paper exploits the global optimisation by graph cut to achieve high quality of matching between images. The disparity map is computed by one cut of the graph with 3D grid topology. To decrease the computational complexity, the estimate of disparity map created by fast local method is used to reduce the disparity space.

2 Acquisition and pre-processing of images

The face capture rig for the passive stereoscopic photogrammetry consists of two digital still cameras and a system of lights. The scheme of capture rig is drawn in Figure 2. Images are taken simultaneously and have resolution 3504×2336 pixels. The purpose of lights is to provide a uniform illumination of the whole face without effects such as shadows or reflections occurring. It is beneficial for the consequent processing and the quality of final model.

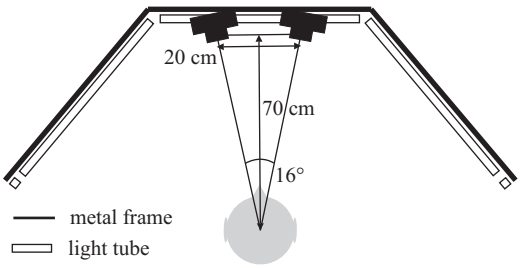


Figure 2: The configuration of face capture rig.

During *a camera calibration* a world coordinate system is established and projection matrices are determined for each camera. The exploited camera model [15] consists of intrinsic and extrinsic parameters. The extrinsic parameters define a relationship between the camera coordinate system and the world coordinate system. The intrinsic parameters describe focal length, sampling of sensor array, position of principal point, radial and tangential lens distortion.

The first stage of image pre-processing is a removal of lens distortion according to the calibration data. The pair of images is then rectified [3]. New rectified projection matrices are found for both cameras. *A rectification* simplifies the traverse along an epipolar line during the correspondence search because it is aligned to pixel row. The last stage is an extraction of face regions in the image pair. The following matching is performed only within these regions. The images are segmented according to a skin model described by Gaussian functions in the RGB colour space. The result of segmentation is closed by binary dilatation and erosion at the end.

3 Construction of disparity space

A matching stereoscopic pair of images means to recognise the pairs of projections of same 3D points. *A matching image* is searched along the epipolar line which corresponds to a processed point in *a reference image* to find a corresponding point. A relationship between matching points can be described by *a disparity*. When input images are rectified, the disparity is simply expressed as a difference of horizontal coordinates of corresponding points. Because of this fact, *a disparity space* can be easily defined as a three-dimensional system with the axes U, V, D. The axes U and V are aligned with rows and columns of the reference image and the axis D represents the disparity. The disparity space has a discrete nature for purposes of this work. Therefore, the integer coordinates $[u_R, v_R]$ address a pixel in the reference image and the disparity d is

measured in whole pixels. A point in the disparity space represents one pair of pixels - $[u_R, v_R]$ in the reference image and $[u_M, v_M]$ in the matching image (Equation 1).

$$u_M = u_R - d \quad v_M = v_R \tag{1}$$

A disparity space image (*DSI*) is a function over disparity space. It defines a measure of confidence that a pair of pixels is corresponding to each other [10]. An assumption that the neighbourhoods of matching pixels in the images are similar is exploited to define this measure. Square windows with fixed size around examined pixels are compared to determine this similarity. The comparison of windows with identical shape implicitly assumes that a projection of same surface patch covers same area in each image (*a window similarity constraint* [14]).

The value $DSI_{u_R,v_R}(d)$ is *a normalised cross-correlation* n_{CC} between the windows centered around the pixel $[u_R, v_R]$ in the reference image and the pixel $[u_M, v_M]$ in the matching image (Figure 3). The pair of pixels is connected by Equation 1. The formulation of n_{CC} is shown in Equations 2, 3, 4 [11].

$$n_{CC} = \frac{cov}{\sqrt{var(R)}\sqrt{var(M)}} \tag{2}$$

$$cov = \sum_{i=-w}^{w}\sum_{j=-w}^{w}(R_{u_R+i,v_R+j} - \bar{R}_{u_R,v_R})(M_{u_M+i,v_M+j} - \bar{M}_{u_M,v_M}) \tag{3}$$

$$var(R) = \sum_{i=-w}^{w}\sum_{j=-w}^{w}(R_{u_R+i,v_R+j} - \bar{R}_{u_R,v_R})^2 \tag{4}$$

The function R_{u_R+i,v_R+j} represents an intensity in the reference image at the pixel $[u_R+i, v_R+j]$ and $\bar{R}_{u_R,v_R}$ is an average intensity over $(2w+1)\times(2w+1)$ window (equivalently for matching image). Higher value of n_{CC} means stronger similarity between compared windows hence bigger confidence that the centres of windows are matching pixels. The n_{CC} is invariant to an affine transformation between intensities in the windows in comparison to SSD or SAD. It brings bigger robustness against reflections or different camera gain. The computation of *DSI* for a compact part of the disparity space can be significantly accelerated [11]. The windows around close pixels are overlapping, therefore many calculations in the n_{CC} are repeated. Equations 3, 4 for a covariance and a variance are rewritten to use only the basic operations between the means of intensity and squared intensity. The means over the overlapping windows arranged in regular array are effectively calculated by box-filtering technique. The speed of the *DSI* computation is then almost invariant with respect to the window size.

A face captured in the stereoscopic image pair has a response inside the *DSI*. It has a form of surface marked by high correlation values. Because a disparity can be seen as an inverse depth, the surface is similar to the real surface of face. Its position in the disparity space is ambiguous in some areas because the course of *DSI* is rugged.

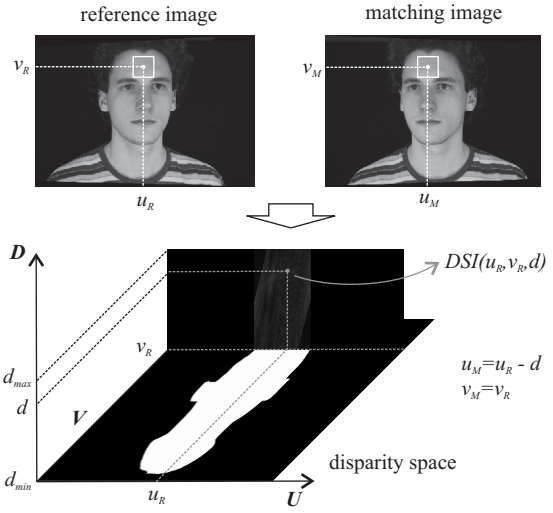


Figure 3: The disparity space with marked entry in the *DSI* for the pair of pixels in images. A slice through the disparity space along row v_R shows that the *DSI* is defined only within the face region.

Main reason for the articulation of *DSI* is that a human skin does not provide strong texture. Secondly, the areas which have view-dependent appearance have problematic matching (in the case of a human face e.g. eyebrows). Finally, the parts which are more oriented to one of the views or completely occluded have an ambiguous response in the *DSI*, because the windows similarity constraint is violated (e.g. sides of nose). The *DSI* is becoming less rugged with increasing size of the correlation window but the present surface is losing details.

The disparity space is large in the case of high-resolution input images, hence it is reduced in several ways in this work. Figure 3 shows that the *DSI* is determined for the pairs of pixels which are inside the face regions in both views. A disparity range is shrunk according to the estimated positions of the nearest and furthest 3D point on the face surface. The reference image is sampled with *a scanning step* and the *DSI* is computed only for these pixels (n_{CC} is calculated from full image resolution).

4 Construction of disparity map

The most complex task in a chain of face capture is an extraction of disparity map from the *DSI*. *A disparity map DM* is a function which assigns to a pixel in the reference image a disparity to its matching pixel. The goal of the correspondence search is to find the *DM* which precisely describes the surface present in the *DSI*. A priori knowledge of reconstructed object gives beneficial constraints on the searched *DM*. A human face has a compact, continuous, and smooth surface. Therefore, *a neighbourhood consistency constraint* [14] can be exploited for the extraction of *DM*. It assumes that neighbouring pixels in a

picture are projections of adjacent spatial points, so their disparity values should be similar. To reflect the reduction of a disparity space, the *DM* is determined only for the sampled pixels within face region in the reference image.

4.1 Local approach

Local techniques determine an entry in the *DM* for particular pixel using its own *correlation function* along the disparity range and optionally the functions of pixels in the neighbourhood. The straightforward method is to assign the disparity value of global maximum in the correlation function to every pixel. However, it leads to the noisy result for a human face because the *DSI* is ambiguous. A concept from [14] was adopted instead but the method is changed to be more reliable for the less-quality *DSI*.

A pixel with strong correspondence can be recognised by a high global maximum in its correlation function and a small ratio between the second highest and highest n_{CC}. Statistics over the face region are gathered to gain general knowledge about these characteristics. The combination of mean and a multiple of variance sets a threshold for each characteristic. *A matching score threshold* t_s and *a ratio threshold* t_r are used to extract the set of pixels with strong matching. The disparity with maximal n_{CC} is recorded into the *DM* for these pixels. There is a possibility that pixels with incorrect disparity (outliers) are added to the *DM* as well because of the statistical approach. The pixels with strong correspondence form a starting point for incremental filling the rest of *DM* using the neighbourhood consistency constraint. The pixels which are on the edge of already resolved area are found in every iteration. An average disparity is computed for each of them from the resolved pixels in 8-pixel neighbourhood. The disparity of the closest local maxima in the correlation function to the average value is recorded to the *DM*. To enforce the smoothness of *DM* and reduce an influence of the outliers in initial set of pixels, *a disparity difference threshold* t_d is established. The disparity of processed pixel is compared to the values of resolved pixels within 8-pixel neighbourhood. The pixel is resolved in certain iteration only if all differences are below t_d in that iteration.

The surface in the *DSI* becomes stronger with enlarging window, especially for the areas of face which are frontoparallel to the camera pair. Consequently, the set of pixels with a strong correspondence covers larger area and contains less outliers. Larger windows (e.g. 31×31 pixels) and t_s, t_r set around mean values lead to better *DM* for the face. The t_d creates 'worm-like' holes which mark rapid depth changes in the *DM*. The local approach is fast but resulting *DM* suffers from imperfections. The slanted parts of surface contain many large regions with incorrect disparity (e.g. the sides of face or nose). Even the areas which are reconstructed better contain steps, although they should be smooth.

4.2 Global approach

Global techniques determine a *DM* in one step using the whole *DSI* simultaneously. Many of them define an optimisation of *DM* as an energy minimisation problem. *An energy function* describes a quality of current shape of the *DM*. A minimum of energy function can be found by various methods such as simulated annealing, mean-field annealing, belief propagation, or *graph cut*. The global approach in this paper is a representative from the family of graph cut techniques. They transform the energy minimisation to *a minimum cut problem* from a graph theory [7]. A minimum cut in a graph is found using *a minimum cut - maximum flow theorem*. The graph cut techniques produce better results for the stereoscopic correspondence than other global methods [10]. Their main disadvantage is, however, higher computational complexity.

The energy function typically consists of two terms [7] as shown in Equation 5. A data term is responsible for choosing the best match for each desired pixel in the reference image. A smoothness term aggregates assumptions about the shape of surface.

$$E(DM) = \sum_{p \in P} DSI'_p(DM_p) + \sum_{(p,q) \in N} \lambda |DM_p - DM_q| \quad (5)$$

P is a set of processed pixels in the reference image and N is a relation between adjacent pixels (4-pixel neighbourhood). λ is *a smoothness coefficient*. DM_p returns a disparity for the pixel p. $DSI'_p(d)$ accesses n_{CC} for the disparity d in the pixel p and transforms it to a 'cost': $c = 1 - ((n_{CC}+1)/2)$ (high correlation means small matching cost). The smoothness term in Equation 5 is linearly dependent on a disparity difference. It enforces everywhere smooth model for the *DM* which is suitable for a human face. The minimisation of this type of energy function can be directly mapped on the computation of one minimum cut in a graph. The 3D grid graph is embedded into the disparity space and the minimum cut is equivalent to searched *DM*. Edge capacities in the graph are set in a way that the cost of minimum cut exactly conforms to the energy function in Equation 5. It is guaranteed that the resulting *DM* corresponds to a global minimum of energy function [12].

The construction of the graph in the disparity space is similar to one proposed in [8]. Figure 4 shows that nodes are created within a disparity range for every desired pixel in the reference image. There is a 6-connectivity between adjacent nodes. Chains of nodes along the axis D are linked by data edges. The capacity of data edge is the matching cost c belonging to the node with higher disparity. When the *DSI* is not defined for some node, the capacity is set to an infinity. One auxiliary layer of nodes is added under the layer with minimal disparity because of the 'shift' of matching costs from nodes to edges. The nodes in each disparity layer are interconnected by the regular 2D grid of smoothness edges with capacity λ. Special edges with infinite capacity connect *a source node* with all nodes in auxiliary layer and all nodes in the layer with

maximal disparity to *a sink node*. Moreover, every data edge has a dual infinite edge oriented in an opposite direction.

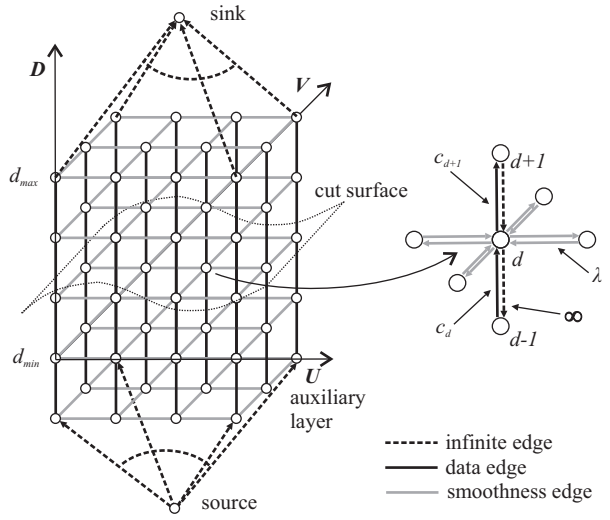


Figure 4: The graph constructed in a disparity space by global approach. The reference image has a resolution 4×3 pixels and the disparity range contains 4 values.

A graph cut can be depicted as a surface which divides nodes into a group with source node and a group with sink node. A set of data edges severed by the cut corresponds to the data term in Equation 5. Identically, an overall capacity of severed smoothness edges is equal to the smoothness term. The linearity of smoothness term is implicitly given by a topology of graph where bigger change of surface is penalised by splitting more smoothness edges. The minimal cut is equivalent to the searched surface in the *DSI*. The resulting *DM* is simply determined by the disparity values corresponding to the severed data edges (one per pixel). A correctness and completeness of *DM* is ensured by the infinity edges which prevent incorrect cuts on the graph. Practically, the minimum cut is found by a maximum flow algorithm because the edges severed by the cut are saturated by a maximum flow. The chosen algorithm is based on a principle of augmenting paths but it is optimised for the graphs with grid topology [2].

The quality of *DM* can be adjusted by the size of correlation window and λ. The fronto-parallel areas are oversmoothed and the slanted areas contain steps in the case of large windows. Decreasing size of window improves the shape of face, the face details appear, but a noise is becoming apparent as well. The increment of λ smooths the surface, however small face details are destroyed. The λ has an influence on a local scale in contrast to the window size. Smaller windows (e.g. 11×11 pixels) and λ set to approximately one eighth of the average matching cost bring the best results for a face. The resulting *DM* describes reasonably the shape of surface including details. The *DM* is smooth with only few outlying regions in strongly slanted areas. It is a significant improvement with respect to the local approach. On the other hand, the constructed graph is huge for high-resolution pictures what implies high memory demands and long computational time.

4.3 Global approach with an estimate by local technique

The high-quality results of global approach are overshadowed by its computational and memory demands. To address this problem, a technique combining the local and global approach was developed. The *DM* produced by the local technique is used for a reduction of search space. Consequently, the global method computes final solution from smaller graph. The comparison with other approaches to the search space reduction such as hierarchical or best-candidate showed that the initialisation by fast local technique brings the biggest memory and time savings [13].

The time of the estimate computation is negligible with respect to following global optimisation. Only significant overhead is computing extra *DSI* using bigger matching windows for better quality of estimate. Initial *DM* is a basis for modelling *a volume of interest* in a disparity space for the global optimisation. The first step is the creation of thin layer with a thickness set by *an offset* o_l around the estimate as it is shown in Figure 5. The disparity range determined by the layer for each pixel is then expanded according to pixel's neighbourhood to eliminate an influence of outliers in the initial *DM*. The minimal boundary of range is updated if lower minimal boundary exists within *an expansion region* (equivalently for the maximal boundary). The expansion region has a square shape with size $(2w_{er}+1) \times (2w_{er}+1)$. Figure 5 depicts a situation when the layer contains correct *DM* after the expansion in spite of the outlier region in the centre of estimate. Finally, the volume of interest is trimmed by the volume where the *DSI* is defined.

The construction of reduced graph starts by creating the 3D grid of primary nodes for the volume of interest as it is depicted in Figure 6 (the grey thin grid in the background shows an extent of full graph). The interconnection scheme is same as in purely global approach except for the edges to the terminal nodes. The chains of primary nodes for adjacent pixels have usually different lengths. It allows an existence of steps in the *DM* that are not penalised by the smoothness term. The solution is to wrap upper and lower boundary of the graph by the auxiliary nodes. They provide ending points for missing smoothness edges. The arrow in Figure 6 marks a situation when valid cut inside the volume of interest is correctly penalised by additional smoothness edge. Data edges between the auxiliary nodes have infinite capacity. The first and the last node of each chain corresponding to a pixel is linked to the source and the sink by infinite edges at the end.

The difficulty with the graph in this form is that the time of graph cut computation is not dependent only on

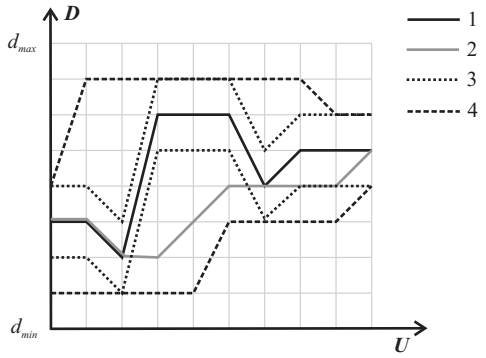


Figure 5: Modelling the volume of interest in a disparity space. A slice along row in the reference image is illustrated (10 pixels and 9 disparity layers). (1) the estimate of *DM* (2) correct *DM* (3) the layer created by $o_l = 1$ (4) the layer expanded by $w_{er} = 2$.

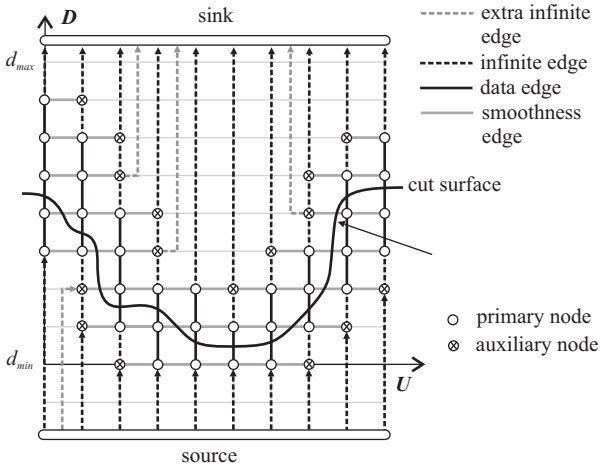


Figure 6: The slice of reduced graph along a row in the reference image (10 pixels and 9 disparity layers). The depiction of edges is simplified except for the edges to terminal nodes.

the number of nodes and edges but also on the shape of graph. When the shape of the volume of interest is distant from a cuboid, the computational time does not linearly decrease together with the size of reduced graph. It is consequence of the optimisation of maximum flow algorithm to regular grid topology of a graph. To cope with this problem, extra infinite edges to the source and the sink are added to the nodes which are on the boundary of graph structure but not in the direction of *D* axis (Figure 6). A flow is quicker brought into main body of graph through them. After this modification, the complexity of graph has smaller influence on a speed of maximum flow algorithm. A solution identical to the one obtained through the purely global technique can be found if the searched *DM* is entirely inside the volume of interest. Otherwise, the boundaries of reduced graph force the minimum cut to globally suboptimal solution.

The effect of window size and λ are the same as in the purely global approach. The size of reduced graph grows together with increasing o_l and w_{er}. The parameters of local technique have to be adjusted to achieve the best estimate of *DM*. The regions with incorrect matching decrease an extent of graph reduction because o_l and w_{er} have to be enlarged to eliminate their influence. Large outliers in the initial *DM* can even prevent the graph cut from finding globally optimal solution because the volume of interest does not contain whole true surface. A tradeoff between the efficiency and the precision of *DM* exists. A price for the solution similar to the purely global approach are high memory and computational demands. However, experiments showed that significant speedup and memory saving with respect to the global technique can be gained with the reasonable loss of precision (only on the sides of face).

5 Surface reconstruction

The last stage of 3D face capture is a transformation of *DM* to a 3D model. An algebraic method involving the projection matrices of each view and the positions of matching pixels is used to compute a 3D point [14]. A triangulation between the 3D points is determined by a neighbourhood of their projections in the *DM*. A simple adaptation of the method in [4] processes independently the square groups of four positions in the *DM*. The level of detail of the model is configured by the scanning step. The resulting model is stored as textured triangle mesh defined in the VRML language.

The surface of model reconstructed by any of presented techniques suffers from local bumpiness which is caused by a noise in the *DSI*. A sub-pixel refinement of *DM* by fitting a parabola to the correlation function [14] does not bring a significant improvement. It operates on smaller scale than the size of present bumps. The *DM* is explicitly smoothed by the Gaussian operator instead. Smoothing enhances visually the quality of surface but the part of small details is lost.

6 Results

The developed technique for stereoscopic matching was evaluated for high-resolution images in terms of a computational time, memory consumption, and quality of reconstructed model. It was compared to the reference local and global technique as well.

Experimental results were obtained using the hardware configuration: Dual Core 3GHz P4 Xeon EM64T, 32GB RAM. The camera calibration was accomplished by external calibration toolbox [1]. The software was imple-

mented in C++ language using Recognition and Vision Library (CVSSP, University of Surrey, UK). Maximum flow computation is performed by external library [2]. The input images showed in Figure 1 were captured under the conditions described in Section 2. A view from the left was chosen as the reference image. The scanning step was 4 pixels and the disparity range was fitted to a face volume (386 layers). The initial estimate by local technique was created with following parameters: the matching window 31×31 pixels, the thresholds t_s and t_r set to the means of corresponding characteristics and the threshold $t_d = 3$ pixels. The consequent global optimisation of *DM* used the *DSI* by matching window 11×11 pixels. The smoothness coefficient λ was equal to 0.025. The volume of interest was modelled by $o_l = 10$ and $w_{er} = 7$. Final smoothing was done by the Gaussian operator with size 13×13 and $\sigma = 3.0$. Equivalent parameters in the reference local and global technique were configured with same values.

Computational times of the image pre-processing were: the optical distortion removal ($8.62s$), the rectification ($26.49s$), the face segmentation ($5m51.06s$). A calculation of n_{CC} for the *DSI* took round $3m15s$ for each technique. A time of surface reconstruction was $10-12m$ depending on the technique. A time of the construction of *DM* itself for each technique is written in Table 1. The time for the global approach with an estimate includes a time for the estimate computation (extra *DSI* + *DM* construction). The used memory listed in Table 1 was almost entirely allocated for the data structures of the *DSI* and the graph. It can be seen that the reduction of graph led to 2.71 times quicker computation and 59.6% reduction of memory use with respect to the purely global technique.

	local	global	global with estimate
time	$5s$	$48m35s$	$17m56s$
memory	$992MB$	$16.96GB$	$6.85GB$

Table 1: The comparison of techniques - the computational time of *DM* construction and the maximal memory consumption during a processing.

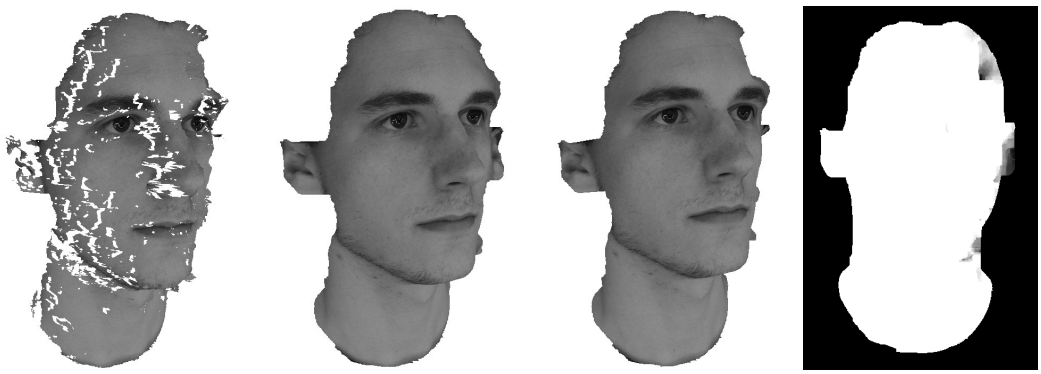
Visual quality of face models reconstructed by individual techniques is illustrated in Figure 7. Although the local technique has the least computational complexity, the model of face contains many holes and outlying patches disconnected from main body of surface. In contrast, the construction of *DM* by a graph cut shows good capability to follow correct surface in ambiguous *DSI*. The result of global technique is a continuous and compact face surface. Fronto-parallel areas are well reconstructed and small details are present (e.g. a shape of eye). A subjective estimate of accuracy in these areas is $\pm 2mm$ from the real shape of face. However, the areas more oriented to one of the views have worse quality (e.g. sides of nose) because of a violation of the window similarity constraint. Occluded parts such as ears are incorrectly reconstructed. The *DM* of the model by local technique was used as a basis for the global approach with an estimate. The degradation of the face model in comparison to the purely global technique is not high with respect to the decrease of computational complexity. Figure 7(d) is a visualisation of absolute difference between the disparity maps by purely global approach and global approach with an estimate. Maximal difference is 207 disparity layers but 95.6% of *DM* is the same. Only the right side of the face is noticeably worse reconstructed because initial estimate contained many strong outliers in this area. in the U

7 Conclusions and future work

In this paper, the hybrid technique merging the local and global approach is proposed for matching the face images from passive stereoscopic setup. Initial estimate of disparity map for the reference image is computed by the fast local technique. The developed local technique gives better results in the case of rugged correlation functions in a disparity space than traditional local methods. The volume of interest is built around the estimate in a disparity space. The global technique based on a graph cut consequently finds optimal solution within this volume. The graph with 3D grid topology models a disparity map as everywhere smooth surface which is determined by single minimum cut. The original algorithm of graph construction from the purely global approach is modified to enable correct and quick computation of a maximum flow in the reduced graph with complex shape.

The proposed technique produces the 3D model of face in good quality which proves that the human skin provides enough details in high-resolution images for precise correspondence search. The central part of the face, which is well visible in both views, is reconstructed on the level of small face details. The sides of the face are not correctly reconstructed because one stereoscopic camera pair with short baseline cannot properly cover whole face. The quality of the model is significantly better than by the local technique. The comparison with the purely global technique shows noticeable degradation of the model caused by the imperfections in the estimate of disparity map. However, a computational time and memory consumption can be significantly smaller with the modest loss of accuracy. A drawback of proposed technique is its dependency on many parameters which have to be adjusted manually for different capture configurations in order to gain the best results.

The main limitation of the technique in terms of model quality is an ambiguity of matching score in a disparity space. Correlation functions in slanted areas could be improved by adaptation of the matching windows with respect to the local orientation of surface. It is also desirable to cope with a noise present in the functions of matching score. Explicit smoothing of disparity map could then be omitted. It is worth to try a hierarchical approach



(a) (b) (c) (d)

Figure 7: Face models reconstructed by local approach (a), global approach (b) and global approach with an estimate by local technique (c). Absolute difference (d) between *DM* from (b) and (c) where the maximal difference is marked by black colour and zero difference by white colour (background is black).

to the construction of disparity map using multiple graph cuts in terms of computational demands. At last, an automatic setup of parameters from general informations about a capture configuration would ease practical use of the method.

8 Acknowledgements

I would like to thank Prof. Hilton (University of Surrey) and doc. Zemčík (Brno University of Technology) for their supervision in different phases of the work. I am also grateful to Dr. Guillemaut and Dr. Starck (University of Surrey) for discussions about graph cuts.

References

[1] J.-Y. Bouguet. Camera calibration toolbox for matlab: www.vision.caltech.edu/bouguetj/calib-doc. Technical report, MRL-INTEL, 2003.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *IEEE Transactions on PAMI*, volume 29, pages 1124–1137, September 2004.

[3] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. In *Machine Vision and Applications*, pages 16–22, March 2000.

[4] A. Hilton, A.J. Stoddart, J. Illingworth, and T. Windeatt. Implicit surface-based geometric fusion. *CVIU*, 69(3):273–291, March 1998.

[5] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *The Fifth European Conference on Computer Vision*, June 1998.

[6] V. Kolmogorov. *Graph based algorithms for scene reconstruction from two and more views*. PhD thesis, Cornell University, January 2004.

[7] N. Paragios, Y. Chen, and O. Faugeras, editors. *Handbook of Mathematical Models in Computer Vision*. Springer, 2006.

[8] S. Roy. Stereo without epipolar lines: A maximum-flow formulation. *International Journal of Computer Vision*, 34(2/3):147–161, 1999.

[9] S. Roy and M.-A. Drouin. Non-uniform hierarchical pyramid stereo for large images. *Vision Modeling and Visualization*, 2002.

[10] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Technical Report MSR-TR-2001-81, Microsoft Research, November 2001.

[11] Ch. Sun. Fast stereo matching using rectangular subregioning and 3d maximum-surface techniques. *International Journal of Computer Vision*, 47(1/2/3):99–117, May 2002.

[12] O. Veksler. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Cornell University, August 1999.

[13] O. Veksler. Reducing search space for stereo correspondence with graph cuts. In *BMVC06*, page II:709, 2006.

[14] I. A. Ypsilos. *Capture and Modelling of 3D Face Dynamics*. PhD thesis, CVSSP, University of Surrey, United Kingdom, September 2004.

[15] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, December 1998.