ITERATIVE VECTOR NON-LINEAR INTERPOLATION FOR THE SYNTHESIS OF INTERMEDIATE VIEWS

Gwenaëlle Le Mestre †, Laurent Lucat † and Danielle Pelé †
France Telecom CNET/DIH † and CNET/DMR †
4 rue du Clos Courtel, BP 59, 35512 Cesson-Sévigné, France
e-mail: danielle.pelle@cnet.francetelecom.fr

ABSTRACT
The purpose of this paper is the synthesis of novel view points from available images captured by a calibrated system. We present the procedure of depth building and back-projection on a virtual image plane that we developed and then mainly focus on the interpolation of missing data in the resulting virtual frame by a multi-spectral non linear filtering technique.

1. INTRODUCTION
The generation of novel views from available viewpoints is a major problem for many applications relative to virtual reality and tele-presence. There are 3 approaches for that purpose : one traditional approach that consists in building a full 3D geometric model of the scene and then rendering it from all viewpoints thanks to many graphical interfaces, a second approach that uses image information and depth, and a third category that uses only images that can be captured with a non calibrated system. The second and last approaches present great interest because they avoid the tedious step of object modelling and can be processed in real time. They are emerging techniques for virtual navigation in a real scene and image based rendering techniques for the visualisation of real or computed generated images [2, 6].

The scope of this paper is related to the second group. The method we propose is a two step procedure, illustrated in Fig 1 for the case of 3 static images calibrated as input. One step is the computation of depth maps from the correspondence analysis of the views and a second step is the back-projection of the computed depth map on the image plane corresponding to a virtual view. The remainder of the paper is structured as follows. Section 2 reviews the method for z (depth) calculation from multiview image analysis that is described in detail in [3] and the principle of the back-projection of z values on virtual image planes. In Section 3, we detail the interpolation method of missing values in the virtual view using a vector non-linear filtering technique.

2. OVERVIEW OF THE ANALYSIS-RECONSTRUCTION SCHEME
We use a calibrated multi-cameras system to acquire views of a real static scene. Then our approach is the following: disparity estimation for each couple of the available stereoscopic pairs of views of the scene, computation and fusion of the depth maps to produce a depth map relative to the virtual camera position. This depth map is then projected on a virtual image plane and post processed with the technique explained in section 3.
2.1. Disparity estimation

Our algorithm is based on correlation on all image pixels since it generates dense disparity maps. The color is used too since it furnishes further information and reduces ambiguities. Because correlation is very sensitive to illumination and colorimetry differences, image illumination and colorimetry are first equalized.

The matching algorithm is applied on a multilevel gaussian pyramid of the images for each color component. The matching process is first applied on the pyramids highest level and transmitted to the next inferior level. The potential matches are initialized by these transmitted values and refined on the searching space. This process is reiterated down to the pyramids last level. The final matching score is a combination of the different color components correlation scores: the MAX value of the three component correlations score is retained. The calculation of the depth map with respect to one camera from the disparity map is then straightforward as the calibration is known.

In the case of more than 2 views, many depth maps can be computed (for example 2 for each camera for the trinocular system). So a fusion criterion taking into account depth coherency, visibility constraints and correlation scores is defined that enables to build a reliable depth map with respect to one original or virtual camera. A label map is associated to the depth map to indicate in which original image to take the luminance and color information depending on the position of the virtual camera relative to the original ones.

2.2. Projection from 3D to virtual 2D image

View point synthesis is achieved by projecting the computed depth map associated to a virtual camera on its corresponding image plane. This calculation is straightforward when the internal parameters of the cameras are known. As the projection will not reach an available sample of the digital image, it is affected to the nearest point of the grid. Projection of depth map and association of the luminance component is not sufficient as problems arise due to inaccuracies in depth estimation of course and image dilation (from discovered areas) and contractions (from recovered areas). Image contractions can be dealt with by the depth fusion criterion that takes into account visibility of the points in space when many depth values correspond to a same pixel. But the dilated zones (the discovered ones) do not receive any value from the projection and so generate holes in the computed image.

So it is necessary to use image interpolation techniques to fill these gaps. We use an original non linear filtering technique that uses RVB components and that is detailed in the next section.

3. INTERPOLATION OF MISSING DATA IN VIRTUAL VIEWS

3.1. Interpolation using a non-linear filter

The aim is, now, to estimate the missing data in the partially reconstructed color image. Numerous techniques are available for data interpolation. In order to choose an adequate processing tool, let us consider the following requirements :

- we want the filter to keep sharp edges, even when the rate of missing data is locally high, as, for instance, in occluded regions
- because missing data occurs very irregularly in the sliding window, the filter is expected to work well with spatial irregular samples.

Non-linear Rank-Order type filters are well suited to these requirements. Such filters involve a preliminary sorting of the input data, and then operate on these sorted samples, no matter if they come from a regular grid or not.
- the filter is then expected to avoid blurring; hence, more importance should be given to samples near the window center, and structured details such as horizontal or vertical lines should be preserved.

This can be achieved using weighted median type filters associated to an adequate set of weights.
- finally, because color images are considered here, the filter should be multi-component in nature, leading to the Weighted Vector Median Filter (WVMF).

The WVMF output is defined as [1]

\[ y_{WVMF} = \arg \min_{x_j} \sum_{i=1}^{N} w_i \| x_j - x_i \|, \]  

where \( \{x_i\}_{i=1,N} \) are the input samples included in the sliding window, \( \{w_i\}_{i=1,N} \) are the associated weights and \( \| \cdot \| \) is the considered norm, usually \( L_1 \) or \( L_2 \). In this work, we used the \( L_1 \) because a fast algorithm is available for the WVMF defined with this norm [4].

When several samples correspond to the \( \min \) in eq. (1), an additional rule is then required to select the output sample.

3.2. Interpolation scheme

The proposed interpolator, which is based on the WVMF, uses a double-iterative structure to reconstruct all missing data. It derives from the interpolation procedure proposed in [5] for mono-valued data. The parameter
which controls the global iteration is the minimal required rate of known samples inside the window, \( \tau_m \).

At the beginning, \( \tau_m \) is set to a high value, e.g. 70%. At each window location, a missing data is thus replaced with the WVMF output only if the rate of available data inside the window (\( \tau \)) is greater than \( \tau_m \).

The whole image is processed several times using the same value of \( \tau_m \) until the interpolator becomes unable to reconstruct enough missing pixels, \( \tau \) being too often lower than \( \tau_m \). This inability is measured through the rate of reconstructed data in the whole image at the given iteration, denoted \( T \). When \( T \) becomes lower than a minimal threshold \( T_m \), say \( T_m = 5\% \), the required rate \( \tau_m \) is decreased, allowing new data interpolations.

Using this double-iterative structure, summarized in Fig. 2, missing data are progressively reconstructed, yielding a good image quality.

3.3. Simplified evolutions of the interpolator

The double-iterative structure presented in §3.2 involves numerous image scannings for the whole data reconstruction, because the required rate \( \tau_m \) of available data inside the window is maintained to a high level. Hence, interpolations are allowed, for a given image scan, only when a sufficient amount of information can be given by the neighborhood of the hole. This leads to a good quality reconstruction at the expense of a high processing time.

When a fast interpolator is needed, one way of reducing the computational complexity is to decrease the rate \( \tau_m \) at each iteration; this leads to a simple-iterative structure which can work significantly faster than the double-iterative one. A major drawback of this approach lies in the critical control of the iteration process: a fast decreasing of \( \tau_m \) should reach too quickly to low rate values, leading to a significant loss of reconstruction quality, and even preventing a total image reconstruction if this rate is not kept at a minimal value.

Another way to speed up the interpolation process, which can be combined with the previous one, is to use a recursive WVMF: when a pixel is reconstructed, it should be considered as available for the next sliding window positions. This also reduces the required number of image scannings. A drawback of this recursivity is that it may produce an image "stretching" which becomes disturbing when the rate of missing data is locally high. This artifact can be nevertheless attenuated using variate image scanning directions.
4. RESULTS

The gaps that are assigned to the 255 value in the synthesized view are reconstructed with our filtering technique. Edge sharpness is well preserved, even in regions containing a lot of missing values. An illustration is given in Fig. 4. Images (a) and (b) form a stereoscopic couple of views of the scene captured with cameras with the same characteristics. We show here a reconstructed virtual view corresponding to a virtual camera with a different focal length (zooming). An interesting part of the partially reconstructed virtual image and its double-iterative WVMP-interpolated version are presented in (c) and (d), respectively, showing the merit of the proposed method. This method has been also applied to the reconstruction of a virtual sequence according to a scenario of virtual visits (including zooming) from the available views.

Acknowledgment

The authors would like to thank Dr. Michael Ropert for fruitful discussions about this work.

5. REFERENCES


Figure 4: Reconstruction and interpolation of a virtual view.