TOF-CCD IMAGE FUSION USING COMPLEX WAVELETS

S. ´A. Gu/GO mundsson and J.R. Sveinsson

Department of Electrical and Computer Engineering
University of Iceland, Reykjavik, Iceland, Email: sag15@hi.is, sveinsso@hi.is

ABSTRACT

A new generation of ToF cameras are practical devices capable of real-time 3D scene reconstruction. They are mainly limited by two factors; accuracy and low spatial resolution. Standard CCD and ToF images of a scene share likenesses in some regions, e.g., object edges are apparent as gradients in both modals. Here we propose an approach to overcome the resolution limitation; using an effective dual-tree complex wavelet transform framework in a calibrated setup to fuse the low resolution TOF image with the high resolution details of the CCD image. We show how this can enhance features such as at object borders.

1. INTRODUCTION

Time of flight (ToF) cameras have proven themselves valuable in computer-vision research [1]. They give instantaneous 3D information of the scene and when added to algorithms increase their robustness. Also, a wide range of new applications in scene acquisition and virtual reality have started to rely on such data. Cameras such as the SwissRanger, PMDtec, Canesta and more are designed as a low-cost solution for such applications. Their depth images are on the other hand low in resolution and contaminated by noise and systematic errors.

In recent years ToF and standard higher resolution CCD images have been fused together using various approaches. The goal of these methods is usually to fuse these somewhat complementary modalities into a high quality CCD intensity/color image with an aligned and enhanced equal resolution depth-map; creating a robust input to various computer-vision applications of interest.

Wavelets are a highly effective multi-scale method that allows us to divide a signal into many scales and subbands where they are analyzed and processed. They have been used in signal processing and fields within computer vision such as compression, feature extraction, image enhancing, registration and image fusion. These methods have been used in multimodal image fusion where the information from the different sensor sources is mixed together in the wavelet domain according to designed fusion rules.

Our proposed fusion method uses a dual-tree complex wavelet transform (DT-CWT) framework to separate the multimodal data individually into coarse and detailed data and then merge the desirable details from the high resolution CCD modality to the low resolution ToF resulting in a high detailed depth-map. Most such 2D/3D fusion approaches exploit the fact that border edges of objects in the two modalities co-align, i.e., at the position of a discontinuity in the depth image there also exists corresponding discontinuity in the intensity image.

In the next section the background theory and related work are introduced. Section 3 presents the algorithm and finally in Section 4 and 5 results and conclusions are given.

2. IMAGE FUSION AND WAVELETS

Multimodal image fusion is a heavily researched area in computer vision, remote sensing, medical imaging robotics, industrial vision etc. The goal is usually to find the complementary aspects of the data available and utilize them to improve the overall data.

2.1. ToF-CCD Image Fusion

Because of the low resolution of ToF cameras and very limited intensity images they produce, some have suggested fusing the ToF data with standard high quality CCD images in a calibrated setup. Huhle et al. [2] projected the TOF 3D points to the standard cameras image plane and then refined these results using a Markov Random Field (MRF) approach. Lindner et al. [3] use a computer graphics approach and a biquadratic filtering for refinement. Bartczak et al. [4] also use a graphics based method for their warping their ToF depth-map. Stereo based fusion methods have also been investigated, e.g., Gu/GO mundsson et al. [5] converted the depth-image into a disparity estimate which was then refined in a hierarchical stereo algorithm using a pair of standard cameras. More recently Zhu et al. [6] used stereo in a dynamic MRF approach where temporal information was used in a globally optimal fashion.

2.2. Wavelet Fusion

Wavelets have been exploited extensively in the image fusion literature. Some are pixel based and others are region based. Some of these methods used standard discrete wavelet transform (DWT), but in recent years other versions and enhancements have been proposed to the wavelets. E.g., Lewis et al. [7] show how a region based DT-CWT fusion algorithm has more directionality than the standard DWT method and is robust to mis-registration due to its shift invariant properties.

2.3. DT-CWT

The dual tree complex wavelet filter bank proposed by N. Kingsbury [8] has been growing in popularity in recent years.

This work was supported by the Research Fund of the University of Iceland.
It is a variant of the standard DWT with properties that make it superior in many applications. Properties such as approximate shift invariance add its robustness in applications such as, motion estimation and image fusion [9],[7]. Also, the directionality of complex wavelets is double, as in each scale level there are 6 directional subbands (Fig. 1). CWT are also free from artifacts such as oscillations (ringing) and aliasing giving better image processing results. Finally DT-CWT are very efficient; the filter bank implementation in the 2D case is $2^2$ redundant which is much less than other shift-invariant DWT. Here we use the DT-CWT Matlab toolbox provided by I.W Selesnick et al. \footnote{http://taco.poly.edu/WaveletSoftware}

3. DT-CWT TOF-CCD IMAGE FUSION.

The general DT-CWT image fusion scheme proposed here is illustrated in Fig. 2. The basic steps are: Wavelet decomposition, merge wavelet coefficients according to the fusion rule and then inverse wavelet transform.

3.1. Registration

We use a Swissranger SR3000 \footnote{http://taco.poly.edu/WaveletSoftware} and a standard CCD monochrome camera that are mounted closely in a rig configuration (Fig. 3). The rig images are registered by a basic projective texturing method based on the multiple CCD-ToF calibration software described by I. Schiller et al. \footnote{http://taco.poly.edu/WaveletSoftware} in [11].

This toolbox calibrates both the intrinsic/extrinsic camera parameters of the rig and corrects the systematic errors of the ToF measurements using a polynomial model. No further processing is made on the missing points due to occlusions and missing singular points as can be seen in Fig. 3.

Instead of projecting the points onto the image plane we build Look up tables (LUT) mapping the points from the ToF image and back for each resolution scale. This way we can wavelet transform the original images and then use the LUTs for the coefficients we want and do not have to worry about the effect of missing points in the wavelet domain.

3.2. Feature Extraction and Fusion Rule

The intensity CCD image has edges due to texture, colors and lighting that we are not interested in for the fusion. Therefore the edges are first detected in the detail wavelet coefficients of the ToF image. The corresponding edge regions are then found in the CDD wavelet coefficients, followed by the fusion. The basic steps of the fusion method are thus:

1. Perform wavelet decomposition of the images: $\omega(I_t) = [A_t, D_t]$ and $\omega(I_c) = [A_c, D_c]$
2. Detect edges in each of the ToF detail subbands $D_{t(l)}$
3. Project the edge region via LUT to the CCD detail subbands $D_{c(l)}$
4. Scale the edge region in $D_{c(l)}$ so its energy is equal to the energy in the edge region in $D_{t(l)}$.
5. Project $A_t$ or $D_{t(l)}$ to the same frame (ToF or CCD) these are now $A_f$ and $D_f(l)$.
6. Take inverse transform to get the edge enhanced depth-map: $\omega^{-1}(A_f, D_f)$.

Where $I_t$ is the 4 times upscaled and extended to $1024 \times 1024$ ToF image, $I_c$ the extended CCD image and $A_e$ and $D_{t(l)}$ are for modality $x$; the low pass approximation and directional subband $\theta$ respectively.

As stated before the images are not registered prior to wavelet transform as the missing points would appear in the wavelet coefficient domain as strong discontinuities and separating this from phenomena from the real edges could prove problematic.

We detect the edge region in the ToF wavelet coefficients by incorporating the denoising method called soft thresholding [12]. We write the soft function as:

$$|d_{t(l)}| = \text{soft}(|d_{t(l)}|, T_t)$$

where

$$\text{soft}(x, T) = \begin{cases} \text{sgn}(x)(|x| - T) & \text{when } |x| > T \\ 0 & \text{otherwise.} \end{cases}$$

Here $T_t$ is calculated for each subband by

$$T_t = \sqrt{\frac{2\sigma_{n,t}^2}{\sigma_m}}$$

where $\sigma_{n,t}^2$ is the “noise” variance at level $l$ and

$$\sigma_m = \begin{cases} \sqrt{\frac{1}{N_\theta} \sum_\theta |d_{t(l)}|^2 - \sigma_{n,t}^2} & \text{when } \frac{1}{N_\theta} \sum_\theta |d_{t(l)}|^2 \geq \sigma_{n,t}^2 \\ 0 & \text{otherwise} \end{cases}$$
\[
\sigma_{n,l}^2 = \alpha_l^2 \frac{1}{N} \sum_\theta |d_t(\theta_l)|^2
\]  
where \(\alpha_l\) is a number smaller than 1 that changes with the decomposition level \(l\). Here this variable was chosen ad hoc as \(\alpha_l = 0.9^l\). \(\sigma_l\) thus always fulfills the upper condition in (4).

The ToF edge region mask is then used to select only the edge region in all CCD subbands. The scale of the CCD image: from 0 to 255 is very different from the ToF image where each pixel denotes a measurement in mm. Therefore a scaling is introduced:

\[
S = \frac{1}{N} \sum_\theta |d_c(\theta_l)|^2
\]

where \(\sum_\theta |d_t(\theta_l)|^2\) is the mean magnitude square of the coefficients or the energy within a subband. In a denoising application the noise variance \(\sigma^2_{n,l}\) would be estimated specifically but here we set it as

\[
\sigma^2_{n,l} = \alpha_l^2 \frac{1}{N} \sum_\theta |d_t(\theta_l)|^2
\]

Finally the inverse wavelet transform is performed, \(I_f = \omega^{-1}(A_l, D_c)\) with \(A_l\) and \(D_c\) in the same frame. Here we choose the ToF frame as it is better for evaluation.

4. RESULTS AND DISCUSSION

Evaluation of data fusion methods is a non-trivial task as often a ground proof is not available or does not apply. The difference in modalities and overall goals of different fusion methods have offspring many evaluation metrics to measure the fusion results. A edge based qualitative test that does not require ground truth images is the Xydeas and Petrovic \(Q^{AB/F}\) measure proposed in [13]. \(Q^{AB/F}\) measures the edge information present, using a Sobel edge detector. A fused image that contains all the input edge information, edge strength and orientation at every pixel, is considered as the ideal fusion result. Here we do not want all the edge information from both images so we modify the metric to our requirements. We manually select the edge region as shown in Fig. 5.

Our modified metric then measures how much of the CCD edge information is present. This is then compared with using just a smooth bilinear upsampled image to measure the improvement. The edge strength problem due to the different modalities is solved by reconstructing the CCD image using the same detail coefficients \(D_c\) as the fusion image and the low pass approximation \(A_c\). This image is shown in Fig. 5 and shows how this reconstruction is sharp on the edge regions and blurred outside of them. In the image we use here and a decomposition level 3 our metric gives 0.35 for our fused result and 0.08 for the bilinear upsampled depth-map.

From Fig. 6 it is evident that the method sometimes leads to some undesirable artifacts, e.g. when edges from different phenomena have fallen inside the edge region. E.g. the sun shines in from the window just above the head causing strong edges in this area that do not belong to the head. These edges cause ripples in the background and thus can often be discarded.

Fig. 6 shows a detailed 3D surface rendition of a small region of the image. In the bilinear version the result is very smooth with a "flying pixel" effect (1) typical for these cameras. In the fused result the edge is clearly much sharper.

5. CONCLUSION

The presented method shows that the modern signal processing method DT-CWT can be used to enhance the details in
Fig. 5. Qualitative evaluation of the fusion algorithm. 
Above: Manually selected edge regions. Inverse transformed 
CCD image using the edge region detail coefficients: \( I_e = \omega^{-1}(A_c, D_f) \) This reconstruction shows how the details outside 
the ToF edge regions have been removed. Below: Edge 
strength of the CCD image. Edge strength of the fused result.

a upscaled depth-map in a simple fusion scheme. The object 
edges are accurately found in the ToF wavelet coefficients and 
can be enhanced using the high resolution details of the corre-
spending CCD coefficients in the calibrated setup. There 

is room for improvement here. E.g. the scaling of the coeffi-
cients due to the dissimilarities of the modalities can done in 
a more advanced and statistically more robust manner than 
just linearly scaling and also the registration is also an area 
we will research in the near future.

6. REFERENCES

[1] Andreas Kolb, Erhardt Barth, and Reinhard Koch, “ToF-
Vision and Pattern Recognition Workshops, pp. 1–6, 2008.
and Edge-Enhanced Distance Refinement for 2D RGB 
[4] Bogumil Bartczak, Ingo Schiller, Christian Beder, and 
Reinhard Koch, “Integration of a time-of-flight camera into a mixed reality system for handling dynamic scenes, 
moving viewpoints and occlusions in real-time,” June 2008.
of stereo vision and time-of-flight imaging for im-
and C. N. Canagarajah, “Pixel- and region-based image 
imaging AG.
ibration of a pmd camera using a planar calibration object 
together with a multi-camera setup.” The Interna-

Fig. 6. Top: Fusion result. Bottom left: Detail from a smooth bilinear interpolation. Bottom right: Same detail 
from the fused result. The fusion edge is clearly sharper.