NOVEL ALGORITHMS FOR OBJECT EXTRACTION USING MULTIPLE CAMERA INPUTS

Jiro Katto and Mutsumi Ohta

Information Technology Research Laboratories
NEC Corporation
1-1, Miyazaki 4-chome, Miyamae-ku
Kawasaki-shi, KANAGAWA 216, JAPAN
E-Mail: katto@dsp.cl.nec.co.jp

ABSTRACT

This paper presents novel algorithms exploiting multiple camera inputs and segmentation techniques, which can be applied to image fusion, disparity detection and object extraction. Differently focused images, stereo pairs and both of them are used for fusion, disparity detection and object extraction, respectively. Firstly, image fusion is done by segmention of each image and determination of focused regions per segment. An efficient decision criterion is developed taking a method of auto-focus into consideration. Secondly, disparity detection is executed by recursively applying segmentation and disparity detection per segment. A new clustering criterion is proposed in order to achieve good segmentation and high compression ratio of disparity maps simultaneously. Finally, object extraction is carried out by utilizing both the fusion result and the disparity map. Experiments are carried out, and they show us effectiveness of the proposed algorithms.

1. INTRODUCTION

Recently, content-based manipulation of photographic video becomes hot topics in multimedia services. The MPEG4 tries to standardize region representation in compressed video bitstreams [1]. The VRML incorporates natural images with region information (alpha maps), which are mapped on geometric objects [2]. A virtual studio is an another example, in which photographic objects are superimposed on a CG background.

Segmentation (object extraction) techniques [3]-[7] play an important role in these applications. The chroma-key is simple but it limits image capturing environment. Single camera approaches do not always work well due to lack of information utilizable. Multiple camera preparation is promising because it provides additional clues to refine segmentation, implies 3D structure recovery, and has robustness against noises.

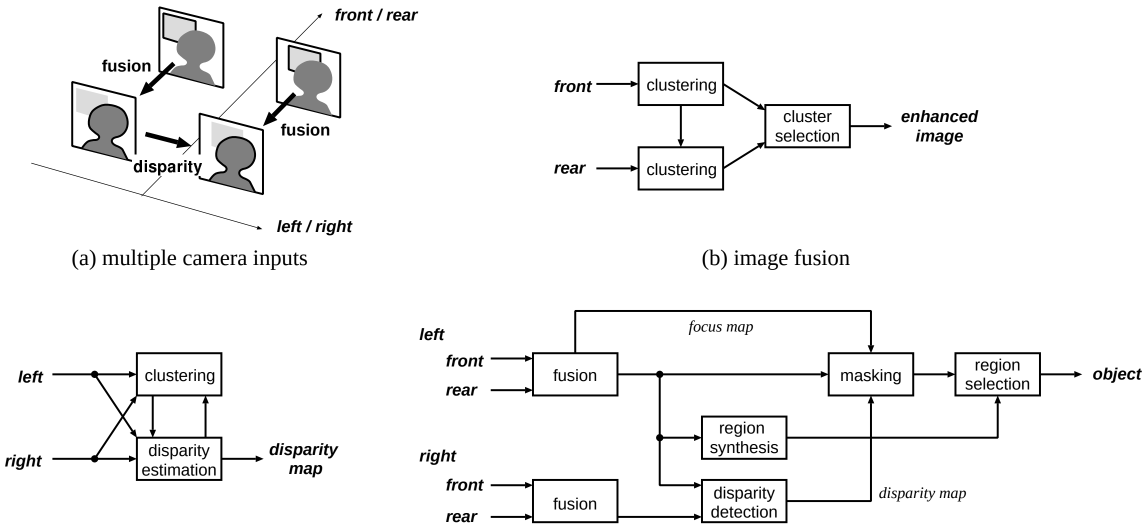
Image fusion [8]-[11] and disparity detection [12]-[14] are related works which utilize multiple camera inputs. The image fusion (in a narrow sense) creates an enhanced image without blurred regions out of differently focused images. The disparity detection attempts to achieve pixel correspondence (disparity) between stereo pairs, leading to structure recovery or efficient compression. Both the results include useful information, focus and disparity, which can be used to improve segmentation performance.

This paper deals with the above issues in a single framework; segmentation techniques are applied to multiple camera inputs. Multiple cameras are located in the same position for image fusion and horizontally different positions for disparity detection. K-means clustering [5, 6, 7] is adopted as the segmentation scheme due to its stability and extensibility. Image fusion algorithm determines focused regions per cluster based on a developed measure taking a method of auto-focus into account. Disparity detection introduces a new clustering criterion, which provides good segmentation results and high compression efficiency of disparity maps simultaneously. Object extraction is finalized using both the results in shrink/expansion strategy. Section 2 describes proposed algorithms, Section 3 presents experimental results, and Section 4 concludes this paper.

2. ALGORITHMS

2.1. Multiple Camera Inputs

Four different cameras are prepared as shown in Fig.1(a). Each of two pairs of cameras shares optical axes. Each pair presents different images; a front object is focused or a rear object is focused. These pairs are located in horizontally different positions.



(a) multiple camera inputs (b) image fusion (c) disparity detection (d) object extraction

Figure 1: Details of the proposed algorithms; (a) multiple camera inputs, (b) image fusion, (c) disparity detection and (d) object extraction.

2.2. Image Fusion

Image fusion is done as shown in Fig.1(b). Firstly, the five-dimensional K-means clustering using (Y, Cr, Cb) components and (x, y) coordinates [5] is applied to each of differently focused images. A distance measure d_n is defined by

$$\begin{aligned} d_n^2 &= w_0 \cdot [(Y - \bar{Y}_n)^2 + (Cr - \bar{Cr}_n)^2 + (Cb - \bar{Cb}_n)^2] \\ &+ w_1 \cdot [(x - \bar{x}_n)^2 + (y - \bar{y}_n)^2] \end{aligned} \tag{1}$$

where $(\bar{Y}_n, \bar{Cr}_n, \bar{Cb}_n)$ is an average value of cluster n, $(\bar{x}_n, \bar{y}_n)$ is a gravity of the cluster, and (w_0, w_1) are weighting factors. Starting from an initial cluster (e.g. block), a cluster minimizing Eq. (1) is searched for each pixel. This operation is repeated in several times, and the clustering is finalized.

Next, by mapping the clustering results to the other image,

$$\sum_{n' \in \mathcal{N}} [(\bar{Y}_n - \bar{Y}_{n'})^2 + (\bar{Cr}_n - \bar{Cr}_{n'})^2 + (\bar{Cb}_n - \bar{Cb}_{n'})^2] \tag{2}$$

is calculated on both images, where $\mathcal{N}$ is a set of clusters neighboring to cluster n. By comparing Eq. (2) between the two corresponding clusters, the cluster with larger value is chosen and an enhanced image is created.

Note that Eq. (2) approximates local variance around the cluster, which some auto-focus techniques also use. An alternative approach, directly comparing variances inside a cluster itself, does not always work well because the clustering gathers smoothed regions in principle.

2.3. Disparity Detection

Disparity detection is executed as illustrated in Fig.1(c). The K-means clustering and disparity estimation per cluster are applied iteratively, and a one-dimensional disparity vector is achieved. A new clustering criterion is defined by extending Eq. (1):

$$\begin{aligned} d_n^2 &= w_0 \cdot [(Y - \bar{Y}_n)^2 + (Cr - \bar{Cr}_n)^2 + (Cb - \bar{Cb}_n)^2] \\ &+ w_1 \cdot [(x - \bar{x}_n)^2 + (y - \bar{y}_n)^2] \\ &+ w_2 \cdot [(Y - Y'_{\vec{v}_n})^2 + (Cr - Cr'_{\vec{v}_n})^2 + (Cb - Cb'_{\vec{v}_n})^2] \end{aligned} \tag{3}$$

where $\vec{v}_n$ is a disparity vector, $(Y'_{\vec{v}_n}, Cr'_{\vec{v}_n}, Cb'_{\vec{v}_n})$ is a pixel value indicated by the $\vec{v}_n$ in the other image, and w_2 is a new weighting factor.

Dissimilar to the previous works, using only five parameters (Y, Cr, Cb, x, y) [5] or appending motion vectors (v_x, v_y)[7], direct utilization of $(Y'_{\vec{v}_n}, Cr'_{\vec{v}_n}, Cb'_{\vec{v}_n})$ leads to drastic reduction of prediction errors. This modification contributes to compression applications; we do not have to encode both the stereo pairs. Instead, only disparity maps are appended to either image as long as efficient disparity maps are obtained.

2.4. Object Extraction

Object extraction is implemented by using both the focus and disparity information in addition to segmentation results. Fig.1(d) depicts the block diagram.

Firstly, image fusion is applied to each of two differently focused images and disparity detection is done between the two fused images. Taking intersection of the focus map (front/rear) and the disparity map (inside/outside the range), an extracting mask is provided. Applying this mask to one of the fused images, initial object extraction is done roughly although some correct clusters are lost. In parallel, region synthesis is done by gathering clusters which have similar statistical properties. The number of masked pixels is then counted in each region determined by the region synthesis. Finally, the regions which contain masked pixels more than pre-determined ratio are selected and compose a final object. This shrink and expansion strategy is analogous to the *opening* in mathematical morphology [15].

3. EXPERIMENTAL RESULTS

Experiments are carried out using real images. Fig.2 demonstrates an image fusion result; (a) and (b) are original and they are fused to (c). Notice that blurred regions are replaced by the other focused ones. Fig.3 gives rate distortion relationship, in which the horizontal axis means bitrates of disparity maps compressed by lossless DPCM, the vertical axis does SNR values of disparity compensated prediction images to the other, and the block refers to conventional block matching. We used three different stereo sequences; originally captured ones and two MPEG sequences (fun_fair and train_and_tunnel). By changing weighting factors, (w_0, w_1, w_2), in Eq. (3), it is recognized that (0,1,1) performs best slightly followed by (1,1,1) but the conventional method, (1,1,0), does not work well from the viewpoint of compression efficiency. As (0,1,1) sometimes results in noisy segments, (1,1,1) seems to be the best choice because it provides good segmentation results and prediction efficiency simultaneously. Fig.4 presents object extraction results; (a) and (b) are obtained by using only focus information or only disparity information, respectively, and (c) is achieved by using both of the clues. Some misjudged regions observed in (a) and (b) disappear in (c) and the extraction performance is definitely improved.

4. CONCLUSIONS

This paper presents efficient algorithms assuming multiple camera inputs. All of image fusion, disparity detection and object extraction are formulated in a consistent manner using clustering techniques. Experimental results convince us of effectiveness of the proposed approaches. Problems, however, lie in depth discontinuity. In case of image fusion, the number of depth indices depends on the number of differently focused images. In case of disparity detection, disparity is uniform inside a cluster. As future works, these problems should be solved.

5. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11/N999: "MPEG4: Testing and Evaluation Procedures Document," (1995).

[2] "The Virtual Reality Modeling Language," Version 1.1 Draft (1995).

[3] R. M. Haralick and L. G. Shapiro: "Survey: Image Segmentation Techniques," CVGIP, pp.100-132 (1985).

[4] M. Kunt, A. Ikonomopoulos and M. Kocher: "Second-Generation Image-Coding Techniques", Proc. IEEE, pp.549-574 (1985).

[5] N. Izumi, H. Morikawa and H. Harashima: "Combining Color and Spatial Information for Segmentation," IEICE Spring Conf., D-680 (1991, in Japanese).

[6] T. N. Pappas: "An Adaptive Clustering Algorithm for Image Segmentation," IEEE Trans. SP, pp.901-914 (Apr.1992).

[7] Y. Yokoyama and Y. Miyamoto: "Image Segmentation for Video Coding Using Motion Information", IEICE Fall Conf., D-150 (1994, in Japanese).

[8] P. J. Burt and R. J. Kolczynski: "Enhanced Image Capture Through Fusion", Proc. 4th ICCV, pp.173-182 (1993).

[9] H. Li, B. S. Manjunath and S. K. Mitra: "Multisensor Image Fusion using the Wavelet Transform", Proc. ICIP'94, pp.51-55 (1994).

[10] K. Kodama, M. Naito, K. Aizawa and M. Hatori: "Enhanced Image Acquisition by Using Differently Focused Multiple Images", ITEC'95, 9-2 (1995, in Japanese).

[11] I. Koren, a. Laine and F. Taylor: "Image Fusion using Steerable Dyadic Wavelet Transform," Proc. ICIP'95, pp.232-235 (1995).

[12] U. R. Dhond and J. K. Aggarwal: "Structure from Stereo - A Review", IEEE Trans. SMC, pp.1489-1510 (1989).

[13] R. Skerjanc and J. Liu: "A Three Camera Approach for Calculating Disparity and Synthesizing Intermediate Pictures," Image Commun., pp.55-64 (1991).

[14] M. Okutomi and T. Kanade: "A Multiple-Baseline Stereo," IEEE Trans. PAMI, pp.353-363 (1993).

[15] R. Haralick, S. R. Sternberg and X. Zhuang: "Image Analysis using Mathematical Morphology", IEEE Trans. PAMI, pp.532-550 (Jul.1987).

(a) front (b) rear (c) integrated

Figure 2: Image fusion results; (a) focused on a front object, (b) focused on a rear object and (c) a fusion result.

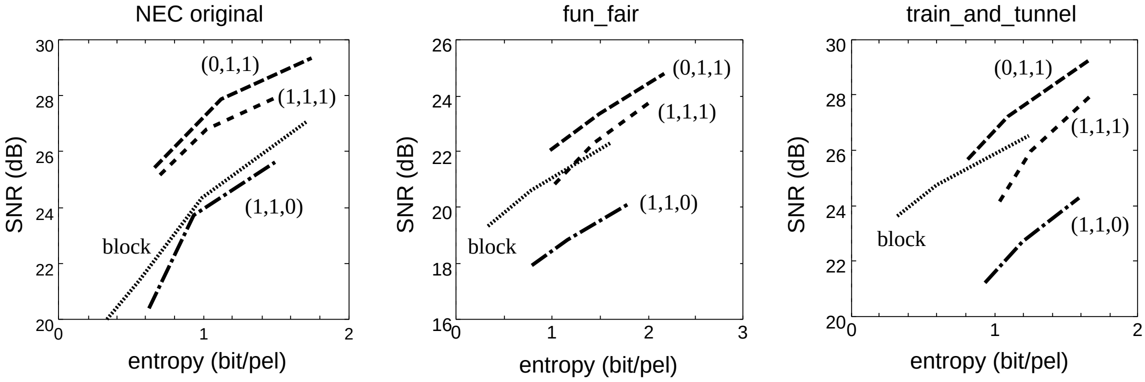


Figure 3: Rate-SNR relationships between disparity maps and disparity compensated predicted images according to weighting factors (w_0, w_1, w_2) in Eq. (3). The horizontal axis represents bitrates of disparity maps compressed by lossless DPCM, and the vertical axis does SNR values of disparity compensated predicted images.

(a) focus information (b) disparity information (c) using both information

Figure 4: Object extraction results according to masking strategy; (a) focus information, (b) disparity information and (c) both of them are used.