Real-time Image-Based Motion Detection Using Color and Structure

Manali Chakraborty and Olac Fuentes

Dept. of Computer Science,
University of Texas at El Paso
500 West University Avenue,
El Paso, TX -79902.
mchakraborty@miners.utep.edu, ofuentes@utep.edu

Abstract. In this paper we propose a method for automating the process of detecting regions of motion in a video sequence in real time. The main idea of this work is to detect motion based on both structure and color. The detection using structure is carried out with the aid of information gathered from the Census Transform computed on gradient images based on Sobel operators. The Census Transform characterizes local intensity patterns in an image region. Color-based detection is done using color histograms, which allow efficient characterization without prior assumptions about color distribution in the scene. The probabilities obtained from the gradient-based Census Transform and from Color Histograms are combined in a robust way to detect the zones of active motion. Experimental results demonstrate the effectiveness of our approach.

1 Introduction

Motion detection is an important problem in computer vision. Motion detectors are commonly used as initial stages in several surveillance-related applications, including people detection, people identification, and activity recognition, to name a few. Changes in illumination, noise, and compression artifacts make motion detection a challenging problem.

In order to achieve robust detection, we combine information from different sources, namely structure and color. The structure information is based on the Census Transform and color information is based on computation of Temporal Color Histograms. We briefly describe both of these techniques later in this section.

1.1 Related Work

The traditional approach for detecting motion consists of building a model of the background using multiple frames and then classifying each pixel in the surveillance frames as either foreground or background. Existing approaches include [10], which proposes spatial distribution of Gaussian (SDG) models, where motion compensation is only approximately extracted, [11], which models each pixel as a mixture of Gaussians and uses an online approximation to update the model and [8], which proposes an online method based on dynamic scene modeling. Recently, some hybrid change detectors have been developed that combine temporal difference imaging and adaptive background estimation to detect regions of change [4]. Huwer et al. [4] proposed a method of combining a temporal difference method with an adaptive background model subtraction scheme to deal with lighting changes. Even though these approaches offer somewhat satisfactory results, much research is still needed to solve the motion detection problem, thus complementary and alternate approaches are worth investigating.

In this paper we propose an approach to motion detection that combines information extracted from color with information extracted from structure, which allows a more accurate classification of foreground and background pixels. We model each pixel using a combination of color histograms and histograms of local patterns derived from the Modified Census Transform [1], applied to gradient images.

1.2 The Census Transform

The Census Transform is a non-parametric summary of local spatial structure. It was originally proposed in [16] in the context of stereo-matching and later extended and applied to face detection in [1]. It has also been used for optical flow estimation [12], motion segmentation [14], and hand posture classification and recognition under varying illumination conditions [6].

The main features of this transform are also known as structure kernels and are used to detect whether a pixel falls under an edge or not. The structure kernels used in this paper are of size 3×3, however, kernels can be of any size $m \times n$. The kernel values are usually stored in binary string format and later converted to decimal values that denote the actual value of the Census Transform.

In order to formalize the above concept, let us define a local spatial neighborhood of the pixel x as $N(x)$, with $x \notin N(x)$. The Census Transform then generates a bit string representing which pixels in *N(x)* have an intensity lower than $I(x)$. The formal definition of the process is as follows: Let a comparison function $\zeta(I(x), I(x'))$ be 1 if $I(x) < I(x')$ and 0 otherwise, let $\otimes$ denote the concatenation operation, then the census transform at x is defined as $C(x) = \bigotimes \zeta(I(x), I(y))$. This process is shown graphically in Figure 1.

1.3 The Modified Census Transform

The Modified Census Transform was introduced as a way to increase the information extracted from a pixel's neighborhood [1]. In the Modified Census Transform , instead of determining bit values from the comparison of neighboring pixels with the central one, the central pixel is considered as part of the neighborhood and each pixel is compared with the average intensity of the neighborhood. Here let $N(x)$ be a local spatial neighborhood of pixel at x, so

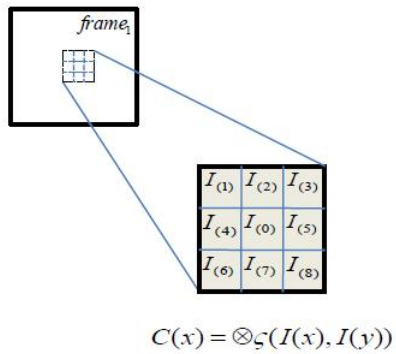

Fig. 1. The intensity $I_(0)$ denotes the value of the central pixel, while the neighboring pixels have intensities $I_{(1)}$,$I_{(2)}$... $I_{(8)}$. The value of the Census Transform is generated by a comparison function between the central pixel and its neighboring pixels.

that $N'(x) = N(x) \cup \{x\}$. The mean intensity of the neighboring pixels is denoted by$I(\bar{x})$. So using this concept we can formally define the Modified Census Transform as follows where all 2^9 kernel values are defined for the 3×3 structure kernels considered.

$C(x) = \bigotimes \zeta(I(\bar{x}),\ I(y))$

1.4 Gradient Image

The Gradient Image is computed from the change in intensity in the image. We used the Sobel operators, which are defined below. The gradient along the vertical direction is given by the matrix G_y and the gradient along the horizontal direction is given by G_x. From G_x and G_y we then generate the gradient magnitude.

$$G_y = \begin{pmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \qquad G_x = \begin{pmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{pmatrix}$$

The gradient magnitude is then given by $G = \sqrt{G_x^2 + G_y^2}$. Our proposed Census Transform is computed from this value of magnitude derived from the gradient images.

1.5 Temporal Color Histogram

The color histogram is a compact representation of color information corresponding to every pixel in the frame. They are flexible constructs that can be built from images in various color spaces, whether RGB, chromaticity or any other color space of any dimension. A histogram of an image is produced first by discretisation of the colors in the image into a number of bins, and counting the number of image pixels in each bin. Color Histograms are used instead of any other sort of color cluster description [3,5,7,9], due to its simplicity, versatility and velocity, needed in tracking applications. Moreover, it has been vastly proven its use in color object recognition by *color indexing* [2,13]. However the major shortcoming of detection using color is that it does not respond to changes to illumination and object motion.

2 Proposed Algorithm

The proposed algorithm mainly consists of two parts; the training procedure and the testing procedure. In the training procedure we construct a look up table using the background information. The background probabilities are computed based on both the Census Transform and Color Histograms. Here we consider a pixel in all background video frames and then identify its gray level value as well as the color intensities corresponding to the RGB color space which are then used to compute the probabilities. In the testing procedure the existing look up is used to retrieve the probabilities from it.

2.1 Training

Modified Census Transform The modified Census Transform generates structure kernels in the 3×3 neighborhood but the kernel values are based on slightly different conditions. To solve the problem of detecting regions of uniform color we use base 3 patterns instead of base 2 patterns. Now, let $N(x)$ be a local spatial neighborhood of the pixel at x so that $N'(x) = N(x) \cup \{x\}$. Then the value of the Modified Census Transform in this algorithm is generated representing those pixels in *N(x)* which have an intensity significantly lower, significantly greater or similar to the mean intensity of the neighboring pixels. This is denoted by $I(\bar{x})$ and the comparison function is defined as:

$$\zeta(I(x), I(\bar{x})) = \begin{cases} 0 & \text{if } I(x) < I(\bar{x}) - \lambda \\ 1 & \text{if } I(x) > I(\bar{x}) + \lambda \\ 2 & \text{otherwise} \end{cases}$$

Where λ is a tolerance constant. Now if $\otimes$ denotes the concatenation operation, the expression for the Modified Census Transform in terms of the above conditions at x becomes;
$C(x) = \bigotimes \zeta(I(\bar{x}), I(y))$

Once the value of the Census Transform kernel has been computed it is used to index the occurrence of that pixel in the look up. So for every one of the pixels we have kernel values which form the indices of the look up table. For a particular index in the look up the contents are the frequencies of pixel occurrences for that value of kernel in all of the frames. When the entire lookup has been constructed from the background video we compute the probabilities which are then used directly at the time of testing.

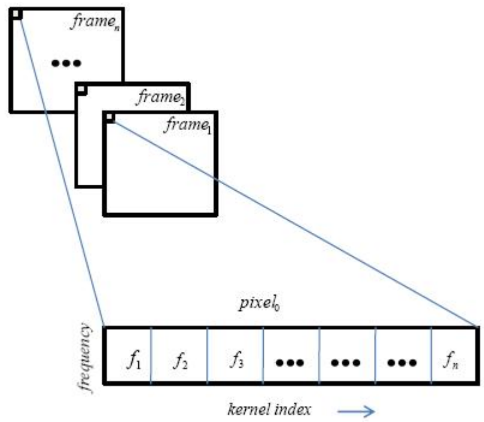

Fig. 2. The frequencies corresponding to the known kernel indices are stored in the given look up for one pixel. The frequency belonging to a particular kernel designates the no of frames for that pixel which has the same kernel value.

Color Histograms The color probabilities are obtained in a similar fashion but using the concept of color histograms. In this case initially the color intensities are obtained in RGB format and based on these values the color cube/bin corresponding to the Red-Green-Blue values of the pixel is determined. Since each color is quantized to 4 levels the sum total of $4^3 = 64$ such bins are possible. So every pixel has a corresponding counter containing 64 values and the color cube values form the indices of this counter look up. The color counter contains the frequency of that pixel having a particular color across all frames.

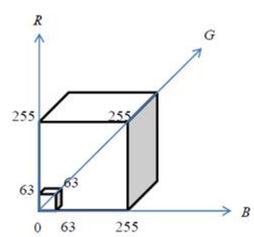

Fig. 3. The color cube displays one of the total $4^3 = 64$ bins possible.

2.2 Testing

The frames from the test video are used to obtain the pixel intensities in both gray level and R-G-B color space. Then for every one of these pixels we compute again the Census Transform which serves as an index in the existing look up for the same pixel and the probability corresponding to that index is retrieved. In this way we use the probabilities corresponding to all pixels in any frame being processed in real time. Fig. 4 demonstrates the concept of testing where initially the Modified Census Transform is computed for pixel_0 based on its 3 $\times$ 3 neighborhood. The kernel value is used as an index for the look up table already constructed to retrieve the required frequency. Here the value of the Census Transform is 25 which then serves as the index. The frequency 33 actually represents that for pixel_0 there are 33 frames which have the same value of kernel of 25.

In the case of color the same scheme is used where every pixel is split into its R-G-B color intensities and then the color bin for it is computed. This is then used as the index to retrieve the value of the probability existing for that particular index from the color counter look up constructed before. So this gives us all corresponding probabilities of pixels belonging to any frame at any instant of time. Once we have the color and the Census Transform probabilities we combine the information from both the color and Census matrices to detect the region of interest.

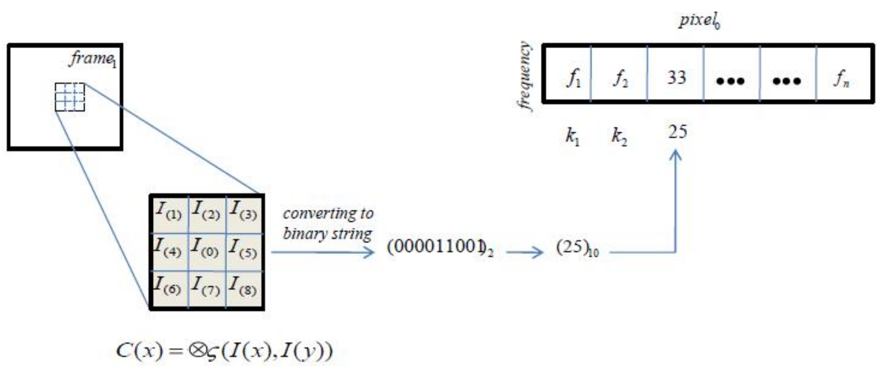

Fig. 4. The Modified Census Transform generates the value 25 which then serves as index for the already constructed look up table during training. The frequency 33 gives the frequency of occurrence of frames for that pixel having the same value of kernel.

3 Results

The first set of experiments was conducted on a set of frames in an indoor environment with uniform lighting and and applying the Census Transform to intensity images. This detection of this set of experiments is shown in Figure 5 and Figure 6. The second set includes frames from videos also taken in an indoor environment, now we used gradient information to compute the Census Transform. The results from this set are displayed in Figure 7. Finally the outdoor training is carried out with a set of frames where there is variation in light intensity as well as fluctuations in wind velocity observed from Figure 8.

Fig. 5. Results from detection using Census Transform probabilities without using gradient information

Fig. 6. Results from detection using color probabilities

4 Conclusion and Future Work

We have presented a method for motion detection that uses color and structure information. Our method relies on two extension to the Census transform to achieve robust detection, namely, the use of gradient information and a base-three encoding. Using gradient information allows to more accurately detect the outlines of moving objects, while our base-three encoding allows to deal effectively with regions of relatively uniform color. The efficiency of the approach

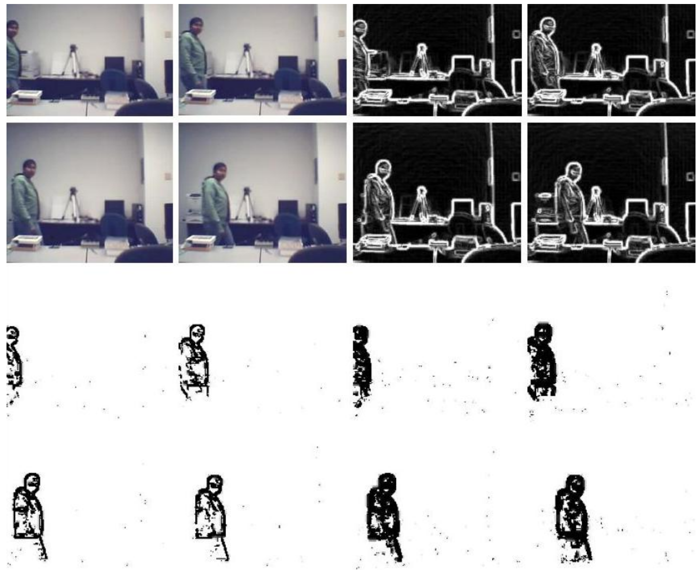

Fig. 7. Results from detection using gradient information to compute Census Transform

is demonstrated by the fact that the entire procedure can be easily carried out in real time. Current work include developing a dynamic update of the background model, developing methods for taking advantage of gradient direction information, and exploring other color spaces.

References

1. Bernhard Froba and Andreas Ernst. Face detection with the modified census transform. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 91– 96, Erlangen, Germany, May 2004.
2. Brian V. Funt and Graham D. Finlayson. Color constant color indexing. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
3. B. Heisele, U. Kressel, and W. Ritter. Tracking non-rigid, moving objects based on color cluster flow. In *Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 257–260, San Juan, Puerto Rico, June 1997.
4. Stefan Huwer and Heinrich Niemann. Adaptive change detection for real-time surveillance applications. In *Third IEEE International Workshop on Visual Surveillance*, pages 37–46, Dublin, Ireland, 2000.
5. Daesik Jang and Hyung-Il Choi. Moving object tracking using active models. *Proceedings of 1998 International Conference on Image Processing (ICIP 98)*, 3:648–652, Oct 1998.
6. Agnes Just, Yann Rodriguez, and Marcel Sebastien. Hand posture classification and recognition using the modified census transform. In *Proceedings of the 7th*

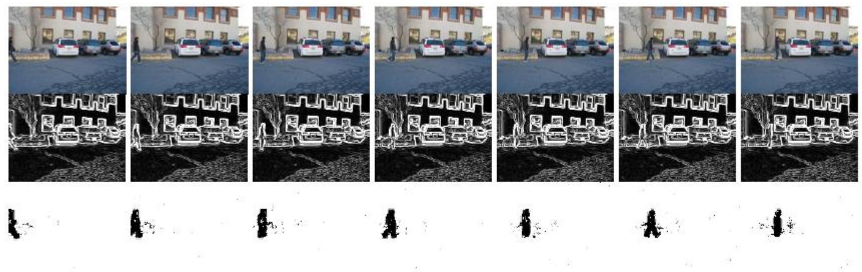

Fig. 8. Results from detection using Census Transform based on gradient images in outdoor environment

International Conference on Automatic Face and Gesture Recognition (FGR06), pages 351–356, 2006.
7. S. J. McKenna, Y. Raja, and S. G. Gong. Tracking colour objects using adaptive mixture models. *Image and Vision Computing*, 17(3/4):225–231, March 1999.
8. Antoine Monnet, Anurag Mittal, Nikos Paragios, and Visvanathan Ramesh. Background modeling and subtraction of dynamic scenes. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 1305–1312, Washington, DC, USA, 2003.
9. T. Nakamura and T. Ogasawara. Online visual learning method for color image segmentation and object tracking. *Proceedings of 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '99)*, 1:222–228 vol.1, 1999.
10. Ying Ren and Ching-Seng Chua. Motion detection with non-stationary background. In *Proceedings of the 11th International Conference on Image Analysis and Processing*, Palermo, Italy, September 2001.
11. Chris Stauffer and W.E.L Grimson. Adaptive background mixture models for real time tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:252, 1999.
12. Fridtjof Stein. Efficient computation of optical flow using the census transform. In *26th Symposium of the German Association for Pattern Recognition (DAGM 2004)*, pages 79–86, Tubingen, August 2004.
13. M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, November 1991.
14. Kunio Yamada, Kenji Mochizuki, Kiyoharu Aizawa, and Takahiro Saito. Motion segmentation with census transform. *Advances in Multimedia Information Processing - PCM 2001*, 2195:903–908, January 2001.
15. N. Yamamoto and K. Matsumoto. Obstruction detector by environmental adaptive background image updating. *ERTICO, editor,4th World Congress on Intelligent Transport Systems*, 4:1–7, October 1997.
16. Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision*, pages 151–158, Stockholm, Sweden, May 1994.