INTEGRAL P-CHANNELS FOR FAST AND ROBUST REGION MATCHING

Alain Pagani *Didier Stricker* *Michael Felsberg*

DFKI, Augmented Vision
Kaiserslautern University, Germany

Computer Vision Laboratory
Linköping University, Sweden

ABSTRACT

We present a new method for matching a region between an input and a query image, based on the P-channel representation of pixel-based image features such as grayscale and color information, local gradient orientation and local spatial coordinates. We introduce the concept of integral P-channels, which conciliates the concepts of P-channel and integral images. Using integral images, the P-channel representation of a given region is extracted with a few arithmetic operations. This enables a fast nearest-neighbor search in all possible target regions. We present extensive experimental results and show that our approach compares favorably to existing methods for region matching such as histograms or region covariance.

Index Terms— Region matching, integral images

1. INTRODUCTION

In this paper, we address the region matching problem, which can be defined as follows: given a reference image and a region of interest, find this region in new images. This computer vision problem is relevant in many tasks, such as feature matching, object detection, recognition, texture classification, 2D and 3D tracking. A central question in region matching is the choice of the region descriptor. For example, SIFT descriptors [1] and affine invariant descriptors [2] are widely used for point matching, but their use for bigger regions with variable shape or size remains cumbersome. In contrast, non-parametric descriptions using e.g. histograms or kernel-density estimates are more adapted to regions. Recently, a new density estimation method called P-channel has been presented [3]. It is a particular type of channel representation [4] that combines the advantages of histograms and local linear models and has been successfully applied in the context of object recognition in [5], where one P-channel representation for the entire image was computed. In this paper, we use the P-channel representation for region matching. To this aim, the P-channel representation of a large number of regions with different scales has to be computed. However, this computation is not tractable if the P-channels are computed in a straightforward way. We therefore introduce a fast method for computing P-channel representations using integral images [6], which has been applied for histograms [7, 8] and region covariance [9].

In this article, we argue that the concept of integral images can also be used for the fast computation of descriptors that include normalized spatial coordinates, like P-channels, for any region in the image. The extraction of the P-channels is fast enough to permit an exhaustive search of the best region in the query image in real time. The main contribution of this paper is our novel algorithm for the fast computation of P-channels using integral images, together with a general method for including normalized spatial coordinates in integral images techniques. Furthermore, we compare integral P-channels with integral histograms and region covariance and show the superiority of the former in a series of experiments.

Section 2 gives a brief overview of the P-channel representation. We then present our approach to region matching using P-channels and derive the fast computation of P-channels using integral images in Section 3. This section also describes the inclusion of spatial coordinates in the integral image technique. We report on our experiments in section 4, and demonstrate the superior performance of P-channels over the standard histogram method and the region covariance method with detailed comparisons.

2. P-CHANNELS

We start by introducing the concept of P-channel representation. Further details can be found in the original paper [3]. P-channel representations, and more generally channel representations [4], are information representations, which can, among other things, be used for representing and estimating distributions of multidimensional feature vectors. In our applications the features are different pixel-based image statistics, such as color, local gradient orientation and local gradient magnitude. However, the concept of channels is more general and can be applied to any kind of multivariate distribution.

2.1. Definitions

Let $\mathbf{f}(\mathbf{x})$ be a pixel based, D-dimensional feature vector. For example, $\mathbf{f}$ can include color and geometric information as follows: $\mathbf{f} = (h, s, |\nabla \mathbf{f}|, \theta)^\intercal$ where $D = 4$, h and s are the

hue and saturation of the pixel, and $|\nabla \mathbf{f}|$ and θ are the gradient magnitude and orientation. We can reasonably assume that fixed bounds can be found for the values obtained in each dimensions. We define the D-dimensional vector of integers $[\mathbf{f}]$ as the component-wise closest integer. An image *region* $\mathbf{R}$ is defined as a set of connected pixels. $|\mathbf{R}|$ is the size (in pixels) of the region $\mathbf{R}$. In this work we focus on rectangular regions, so that $\mathbf{R}$ can be defined by an upper-left pixel $\mathbf{x}_0 = (x_0, y_0)$ and a lower-right pixel $\mathbf{x}_1 = (x_1, y_1)$, and $|\mathbf{R}| = (x_1 - x_0)(y_1 - y_0)$. A region *descriptor* is a representation of the distribution of the vectors $\{\mathbf{f}(\mathbf{x}), (\mathbf{x}) \in \mathbf{R}\}$, in the feature-space. For example, this distribution can be represented by the sample mean vector $\bar{\mathbf{f}}$, the sample covariance matrix, or through a D-dimensional histogram.

2.2. Histogram vs. P-channels representation

A histogram is actually the most trivial case of a channel representation. Without loss of generality, the bins have unit size in every dimension, and the bin centers are located at integer positions of the D-dimensional feature space (this amounts only to scaling in each dimension). Thus, the location of a bin is represented by a multi-index $\mathbf{i}$, i.e. a vector of indices $(i_1, \ldots, i_D)$. If the histogram uses n bins in each dimension, the feature space is tesselated into n^D channels (the bins), each of them having a projection operator of the following form:

$$\mathbf{h_i}(\mathbf{R}) = \frac{1}{|\mathbf{R}|} \sum_{\mathbf{x} \in \mathbf{R}, [\mathbf{f}(\mathbf{x})] = \mathbf{i}} \left(1 \right) \tag{1}$$

Equation (1) is the classical normalized histogram encoding function, where each bin $\mathbf{h_i}$ stores a count of the feature vectors falling in that bin (provided by the test function $[\mathbf{f}(\mathbf{x})] = \mathbf{i}$), divided by the total number of feature vectors $|\mathbf{R}|$. Each bin stores one real number, and the complete histogram can be stored using n^D values.

The P-channel representation can be seen as an extension of histograms, where each bin (each channel) stores as additional information the sum of offsets from the channel center. Each channel thus stores a $(D+1)$-dimensional vector constructed as follows:

$$\mathbf{p_i}(\mathbf{R}) = \frac{1}{|\mathbf{R}|} \sum_{\mathbf{x} \in \mathbf{R}, [\mathbf{f}(\mathbf{x})] = \mathbf{i}} \begin{pmatrix} \mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{x})] \\ 1 \end{pmatrix} \tag{2}$$

The complete P-channel representation can be stored using $n^D \times (D+1)$ values.

2.3. P-channels with spatial features

As mentionned before, the feature vectors $\mathbf{f}(\mathbf{x})$ usually include several pixel statistics such as color, gradient orientation and magnitude. A specificity of the P-channel representation is that the feature vector is completed with the explicit normalized spatial coordiates of the pixel (x, y) (normalization over the considered region). As a result, the feature space is extended to $(D+2)$ dimensions. If we use n_x, resp. n_y bins in the x, resp. y dimensions, we end up with a P-channel representation with $n^D \times n_x \times n_y$ channels, and each P-channel contains a $(D+3)$-dimensional vector. The complete representation can be stored using $n^D \times n_x \times n_y \times (D+3)$ values. Equation (2) remains valid, if we consider $\mathbf{f}(\mathbf{x})$ as a $(D+2)$-dimensional vector (the two last one being the normalized spatial coordinates), and $\mathbf{i}$ a $(D+2)$-dimensional vector of indices.

In order to compare two P-channel representations, we use the Euclidean distance. Although P-channel representations are not vectors, the Euclidean distance is sufficient for a robust matching and has the advantage of beeing extremely fast.

3. REGION MATCHING USING INTEGRAL P-CHANNELS

The task we propose to solve is the following: given a reference image and an object of interest (e.g. a car, a face, a building), we define a region surrounding this object. The problem is to find the same (or an appropriate) region in subsequent images. For this, we have to perform an exhaustive search for all regions and scales in the query image. However, computing the P-channels individually for all possible region centers and a reasonable number of scales is not feasible in near real time. An exhaustive search on 70000 regions covering 19 scales takes more than 15 seconds for a 320 by 240 image. However, this time consumption can be significantly reduced (a few hundred milliseconds for the same image size and same number of regions) if we use the *integral image* [6] representation.

The idea of the integral P-channel formulation is that we can globally compute an integral P-channel representation for the entire image in a preprocessing step, and deduce the P-channel representation of a given region from the integral P-channel in a few arithmetic operations.

However, the P-channel representation uses normalized spatial coordinates: the spatial coordinates are scaled and shifted so that the width and the height of the region always range from 0 to 1. This step is necessary to allow scale and translation invariance when comparing the P-channel representation of two different regions. Thus it is impossible to construct directly an integral image for P-channels with spatial coordinates. We solve this problem by introducing an intermediate P-channel $\mathbf{q_i}$, the encoding of which is defined as follows:

$$\mathbf{q_i}(\mathbf{R}) = \sum_{\mathbf{x} \in \mathbf{R}, [\mathbf{f}(\mathbf{x})] = \mathbf{i}} \begin{pmatrix} \mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{x})] \\ x_{abs} \\ y_{abs} \\ 1 \end{pmatrix} \tag{3}$$

Note that in equation (3), x_{abs} and y_{abs} are the absolute spatial coordinates in the image, $\mathbf{f}(\mathbf{x})$ is a D-dimensional vector, and $\mathbf{i}$ a D-dimensional vector of indices. The complete representation $\mathbf{q}$ is stored using $n^D \times (D+3)$ values.

The integral P-channel representation is defined as the intermediate P-channel representation of the region $\mathbf{R}(0,0,x,y)$ between the origin and the point (x,y):

$$\mathbf{Iq_i}(x,y) = \mathbf{q_i}(\mathbf{R}(0,0,x,y)) \tag{4}$$

Like other integral image techniques, the integral P-channel can be computed incrementally for every pixel of the image in one single pass.

Once the integral P-channel $\mathbf{Iq}$ has been generated, it is possible to compute the P-channels $\mathbf{q_i}(\mathbf{R}(x_0,y_0,x_1,y_1))$ for any target region with a few arithmetic operations:

$$\begin{aligned}\mathbf{q_i}(\mathbf{R}(x_0,y_0,x_1,y_1)) \quad = \quad & \mathbf{Iq_i}(x_1,y_1) - \mathbf{Iq_i}(x_1,y_0) \\ & -\mathbf{Iq_i}(x_0,y_1) + \mathbf{Iq_i}(x_0,y_0)\end{aligned} \tag{5}$$

We now show how to construct the P-channels $\mathbf{p_i}(\mathbf{R})$ (with normalized spatial coordinates) from $\mathbf{q_i}(\mathbf{R})$. The region $\mathbf{R}$ is first tesselated into $n_x \times n_y$ cells $C_{j,k}, (j,k) \in [1 \ldots n_x] \times [1 \ldots n_y]$. For each cell, an intermediate P-channel $\mathbf{q}_{\mathbf{i},C_{j,k}}$ is extracted from the integral P-channel. The P-channels $\mathbf{p}_{(\mathbf{i},j,k)}$ are then normalized from the P-channels $\mathbf{q}_{\mathbf{i},C_{j,k}}$, as follows:

$$\begin{aligned} p^d_{(\mathbf{i},j,k)} &= \tfrac{1}{|\mathbf{R}|} q^d_{\mathbf{i},C_{j,k}} \qquad \textit{if } d \in [1,D] \cup \{D+3\} \\ p^{D+1}_{(\mathbf{i},j,k)} &= \tfrac{1}{|\mathbf{R}|(u_1-u_0)} (q^{D+1}_{\mathbf{i},C_{j,k}} - \tfrac{u_0+u_1}{2} q^{D+3}_{\mathbf{i},C_{j,k}}) \\ p^{D+2}_{(\mathbf{i},j,k)} &= \tfrac{1}{|\mathbf{R}|(v_1-v_0)} (q^{D+1}_{\mathbf{i},C_{j,k}} - \tfrac{v_0+v_1}{2} q^{D+3}_{\mathbf{i},C_{j,k}}) \end{aligned} \tag{6}$$

where (u_0,v_0) and (u_1,v_1) are the coordinates of upper-left and lower-right corners of the cell $C_{j,k}$.

4. EXPERIMENTS

We compare the performance of our P-channel region matching with integral histograms [7] and region covariance matching [9]. For the integral P-channel method and the integral histogram method, we used color images, and the features are the pixel's hue, saturation and the local orientation of the gradient (computed on a grayscale image). For the P-channels, the number of bins is $n = 3$, and spatial dimensions are split into $n_x = n_y = 2$ bins, resulting in the computation of 162 integral images. For the histograms, the number of bins is $n = 5$, resulting in the computation of 125 integral images. For both methods, we used the Euclidean distance between representations. For the region covariance method, we used the 9-dimensional feature vector: pixel location (x,y), color values (RGB), and the norm of the first and second order derivatives of the intensities, resulting in the computation of 54 integral images. The distance between covariance matrices is the Förstner distance [10]. For the three methods, we apply the matching refinement method of [9]. For a number of video sequences, we take one image as reference image and manually define a reference region. We then perform an exhaustive search in the remaining images for a number of query regions. The size of the images is 320 by 240 pixels. We search for 19 different scales (15% scaling between consecutive scales) at location centers every 6th pixel. For a reference region of 100 by 100 pixels, this results in approximatively 70.000 search regions for a single frame, which are tested in a fraction of a second when using integral images.

Fig. 1. Comparison of three methods for region matching. Left: reference image and region. Middle and right: output of the three methods for different images (see text for details).

4.1. Sequences

We tested the three methods with 4 different sequences. Figure 1 shows some frames of the video sequences with the reference image and region in the first columns and the bounding boxes found by the three methods in the remaining columns (red: integral P-channels, blue: integral histograms, green: region covariance). The first row shows the application of our method to a semi-planar patch with partial pose change. Our method finds the right region even with large scale variations and motion blur. The second row shows the results with variations of the incident light on the object's surface. The third

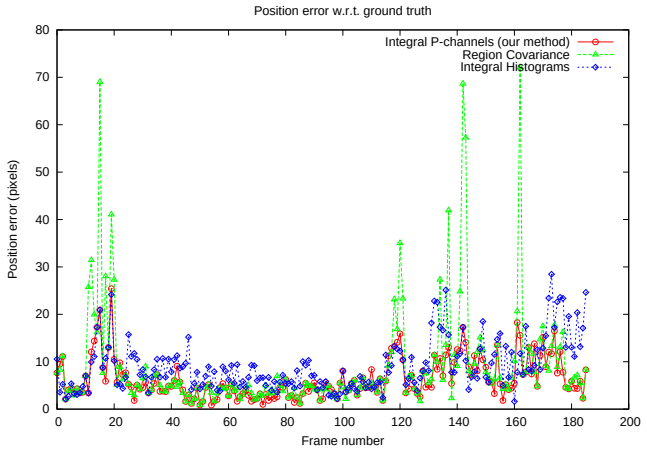

Fig. 2. Error w. r. t. manually marked ground truth

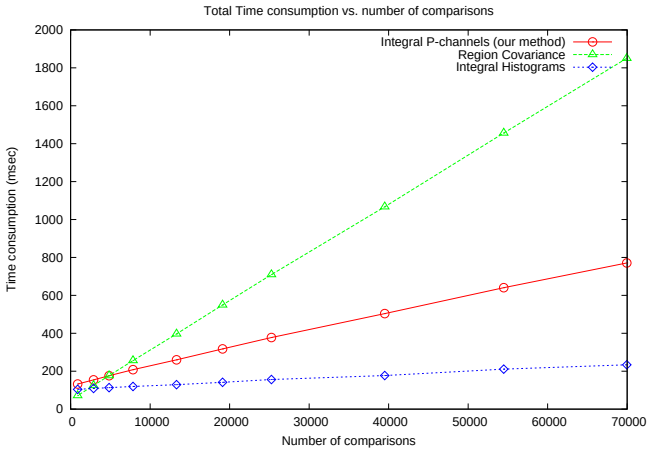

Fig. 3. Time consumption of three methods.

and fourth row show the performance of our region matching algorithm with a 3D object. Note the robustness to scale changes and camera position.

Figure 2 shows the average error on the corner positions with respect to a manually marked ground truth for another sequence. Our P-channel method shows a strong robustness to motion blur (around frame 10 to 20, 120, 140 and 160) in comparison with region covariance. In all the sequences, our method shows slightly better results than the region covariance method, and much better results than the integral histograms method. This can be explained by the fact that the integral histograms do not use the spatial information, as our method and the region covariance do.

4.2. Time consumption

We compared the time consumption of the three methods for varying number of candidate regions (see Figure 3). The results show that while our approach is slower than the integral histograms, it is faster than the region covariance and yields slightly better results. In comparison with region covariance, the P-channel method is 1.5 times faster for a usual set of 13.000 regions, and is even more attractive when the number of regions increases.

5. CONCLUSION

In this paper, we introduced a novel method for the fast computation of the P-channel representation of a large number of regions in an image using the approach of the integral images. Moreover, we added our contribution to the integral image technique by showing how to compute integral images for descriptors including normalized spatial coordinates. The comparison with other region matching techniques showed slightly better results than region covariance while being $1,5$ times faster. This novel technique opens new possibilites for the P-channel desciptor, which could be used as a generalization of histograms in many image processing methods. For our future work, we will consider using the P-channel representation and local integral images for fast region tracking between successive frames.

Acknowledgment - The research leading to these results has been partially funded by the German BMBF project AVILUSplus (01IM08002) and by the European Communitys Seventh Framework Programme (FP7/2007-2013) under grant agreement Nr 215078 DIPLECS.

6. REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Comp. Vision*, vol. 60, pp. 91–110, 2004.

[2] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comp. Vision*, vol. 60, pp. 63–86, 2004.

[3] M. Felsberg and G. Granlund, "P-channels: Robust multivariate m-estimation of large datasets," in *ICPR*, 2006.

[4] G.H. Granlund, "An associative perception-action structure using a localized space variant information representation," in *AFPAC*, 2000.

[5] M. Felsberg and J. Hedborg, "Real-time visual recognition of objects and scenes using P-channel matching," in *SCIA*, 2007.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001.

[7] F. Porikli, "Integral histogram: a fast way to extract histograms in cartesian spaces," in *CVPR*, 2005.

[8] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *CVPR*, 2006.

[9] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *ECCV*, 2006.

[10] W. Förstner and B. Moonen, "A metric for covariance matrices," *Quo vadis geodesia*, pp. 113–128, 1999.