A NEW HYBRID DESCRIPTOR BASED ON SPATIOGRAM AND REGION COVARIANCE DESCRIPTOR

Niloufar Salehi Dastjerdi and M. Omair Ahmad

Department of Electrical and Computer Engineering,
Concordia University, Montreal, Quebec, Canada

ABSTRACT

Image descriptors play an important role in any computer vision system e.g. object recognition and tracking. Effective representation of an image is challenging due to significant appearance changes, viewpoint shifts, lighting variations and varied object poses. These challenges have led to the development of several features and their representations. Spatiogram and region covariance are two excellent image descriptors which are widely used in the field of computer vision. Spatiogram is a generalization of the histogram and contains some moments upon the coordinates of the pixels corresponding to each bin. Spatiogram captures richer appearance information as it computes not only information about the range of the function like histograms, also information about the (spatial) domain.However, there is a drawback that multi modal spatial patterns cannot be well modelled.Region covariance descriptor provides a compact and natural way of fusing different visual features inside a region of interest. However, it is based on a global distribution of pixel features inside a region and loses the local structure.In this paper, we aim toovercome the existing drawbacks of these descriptors. To this, we propose r-spatiogram and then a new hybrid descriptor is presented which is combination of r-spatiogram and traditional region covariance descriptors. The results show that our descriptors have the discriminative capability improved in comparison with other descriptors.

KEYWORDS

Feature Descriptor, Spatiogram, Region Covariance

1. INTRODUCTION

Feature descriptors is one of the main building blocks in any computer vision system e.g. object recognition and tracking. It is utilized for comparing image or images patches so appropriate description of features in the image is the desired goal in such applications. Many approaches for the feature description are proposed such as Gaussian Mixture Models (GMMs) [1] and region covariance descriptor [2]. In GMM, pixels are grouped into homogeneous region by grouping the signal vectors in a selected dimensional space. The use of a GMM always provokes questions regarding the number of mixtures to use. Region covariance descriptor describes a region of interest as a covariance of pixel features. It provides a natural way to fuse different modalities, e.g., color and texture, of pixel features into a single meta-descriptor. Since the covariance descriptor is obtained by averaging features inside the region, it remedies the effects of noise and spatial misalignments.However, Covariance descriptor is based on a global distribution of pixel features inside a region and loses the local structure. Hence, two different patterns that have

the same distribution of pixel features, have similar covariance descriptors Moreover, many approaches used to form a description involve the use of a template or a histogram. Lowe [3] suggested scale-invariant feature transform(SIFT) algorithm which has a local feature detector and local histogram-based descriptor. It detects sets of interest points in an image and for each point it computes a histogram-based descriptor with 128 values. Sedaghat and Ebadi [4] proposed a new local feature descriptor that is named adaptive binning SIFT (AB-SIFT) by exploiting an adaptive binning approach to describe the image content around a local feature and utilizing an adaptive histogram quantization strategy for both the location and gradient orientations respectively, that increases the discriminability of the final descriptor. Farenzena et al. [5] presented the famous SDALF method which combined weighted HSV histogram with stable color region and salient texture. The SCEAF [6] method tried to improve SDALF with detected human parts or structural constraints. Ma et al. proposed eBiCov [7] descriptor based on fisher vectors and bioinspired features.Su et al. [8] presented the histogram of oriented gradient (HOG) descriptor that applied weight for every bin of gradient orientation histogram according to the significance of the gradient information. Considering that histogram discards all spatial information, Birchfield and Rangarajan [9] introduced spatiogram. Spatiogram is a generalization of the histogram and contains some moments upon the coordinates of the pixels corresponding to each bin. The spatiogram computes not only information about the range of the function like histograms, but also information about the (spatial) domain. However, one drawback of representing the spatial information bymeans of a single Gaussian is the fact that multi modal spatialpatterns cannot be well modelled. So, this approach sometimes may consider two objects with the different distribution as the same object.Nilsson et al. [10] proposed mapogram which incorporates spatial information into a histogram. In mapogam, the spatial information is contained in 2D probabilistic "maps". It is possible to control the amount of spatial information desired in the model. Mapogram may require the great amount of memory for computing the map and need to scan overall image iteratively. Choe et al. [11] introduced geogram that contains information about the perimeter of grouped regions in addition to features in the spatiogram. However, for some images, the mentioned drawback of spatiogram still remains in geogram approach. So, in order to alleviate the mentioned drawbacks in region covariance descriptor and spatiogram we summarize our main contributions as follows:

- To overcome the drawback of spatiogram, r-spatiogram is introduced based on the region-based compactness on the distribution of the given feature. In addition, a modification to the similarity measure for a spatiogram[12] is respectively proposed in this paper.
- Targeting the more complete representation of image, a new descriptor, r-spatiocovariance is proposed which is combination of r-spatiogram and traditional region covariance descriptor and includes more rich features of the images.
- The rest of the paper is organized as follows. Section 2 describes the details of the proposed method. . The experimental results and performance evaluations are provided in Section 3. Finally, the conclusion are drawn in Section 4.

2. PROPOSED METHOD

The proposed method is constructed to provide proper and applicablerepresentation of images which applied useful feature components and descriptors. Since spatiogram and covariance descriptors have been acknowledged and widely used for their excellent representation abilities, they are utilized in this work.

2.1. r-spatiogram

Spatiogram is often calculated on the whole image region, which lacks robustness as a global image region is not discriminative enough to describe a shape. Hence, given an image that is a two-dimensional mapping $I: P \rightarrow v$ from pixels $P = [x, y]$ to values v, we divide the image region, to several subregions. Then, for each bin, b, and in each subregion, r, we compute the ratio of the each subregion's histogram, $r_i n_b$, to whole image's histogram, In_b, as (1).

$$\boldsymbol{r_{ib}} = \frac{r_i n_b}{In_b}, \qquad b = 1, \dots, B; i = 1, \dots, M \tag{1}$$

where n_b denotes the number of pixels whose value is that of the bth bin, B is the number of bins in the spatiogram, and M shows the number of subregions in the image. After vectorizing the computed ratios in (1), we obtain $\boldsymbol{r_b}$ which presents a vector of ratios for corresponding bin with the length equal to the number of subregions. Finally, the third order r-spatiogram of an image, rs, is represented as,

$$rs_b^3 = \langle n_b, \mu_b, \Sigma_{b,} r_b \rangle, \qquad b = 1, \dots, B \tag{2}$$

where μ_b and Σ_b are the mean vectors and the covariance matrices of the coordinates of those pixels, respectively. The motivation we introduced ratios stems from considering region-based compactness on the distribution of the given feature in the image. In order to obtain the similarity between two r-spatiogram, first we use l1-norm to calculate the similarity, s, between ratios of two image, $\boldsymbol{r'_b}$ and $\boldsymbol{r'_b}$, according to (3). Then, the final similarity measure, ρ, which introduced in [12] is modified as (4),

$$rDist = \|\boldsymbol{r_b} - \boldsymbol{r'_b}\|_1 = \sum |r_{ib} - r_{ib'}|,$$
$$s = |1 - rDist| \tag{3}$$

$$\rho = \sum_{b=1}^{B} s \sqrt{n_b n'_b} \left[8\pi |\Sigma_b \Sigma'_b|^{1/4} N(\mu_b; \mu'_b, 2(\Sigma_b + \Sigma'_b)) \right] \tag{4}$$

To analyze the obtained measure in (4), we compare two identical spatiograms and compute the total similarity as (5). As can be seen, any spatiogram compared to itself will always receive a similarity score of 1 using the new measure, which is its maximal value we obtain,

$$\rho = \sum_{b=1}^{B} \sqrt{n_b^2} \left[8\pi |\Sigma_b|^{1/2} N(\mu_b; \mu_b, 4\Sigma_b) \right]$$
$$= \sum_{b=1}^{B} 8\pi n_b \frac{|\Sigma_b|^{1/2}}{2\pi |4\Sigma_b|^{(1/2)}} = \sum_{b=1}^{B} n_b = 1 \tag{5}$$

2.2. r-spatiocovariance

Covariance descriptor describes a region of interest as a covariance of pixel features and fuse different modalities such as color and texture of pixel features into a single meta-descriptor. However, covariance descriptor is based on a global distribution of pixel features inside a region and similar to other meta-descriptors loses the local structure. Figure1(a) and Figure1(b) indicate two different patterns that have the same distribution (mean/covariance) of pixel features (each color indicates the same feature vector)[13]. So, they have similar

covariance descriptors that can be seen in Figure1(c) and Figure1(d). Note that the features used for this figure are red, green and blue channel of color image.
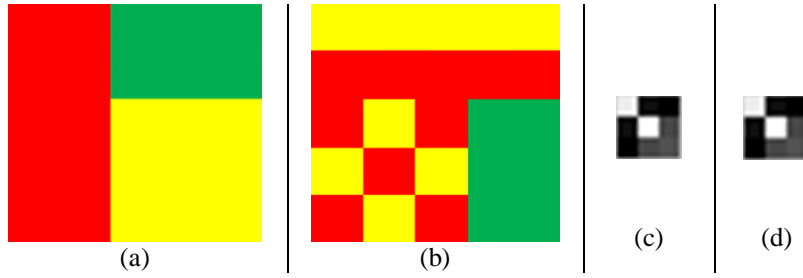

Figure 1. (a) First pattern, (b) Second pattern, (c) Covariance descriptor for first pattern, (d) Covariance descriptor for second pattern.

In this paper, making use of mean information to enhance the covariance descriptor is motivated by several works, such as [14, 15, 16]. Inspiring from [16] which incorporated the first-order statistics to the representation of region covariance, the covariance descriptor is computed as (6).

$$R(c) = (\mu, s_1, \dots s_d, s_{d+1}, \dots, s_{2d})^{\mathrm{T}}, \tag{6}$$

where the first term, μ, is the mean vector of the features and the remaining terms denote a vectorial representation of a covariance matrix by concatenating the elements of Sigma Points $S = \{s_i\}, i = 1, \dots, d$ for a $d \times d$ covariance descriptor. More details about sigma point is in [17] To compute the distance between two feature vectors, $R(c)$ and $R'(c)$, l1-norm is utilized as (7),

$$\beta = \|R(c) - R'(c)\|_1 \tag{7}$$

Finally, weighted combination of two feature components, computed in (4) and (7), is represent as (8) which calculates the distance between two images.

$$D(I, I') = w_\beta \beta + w_\rho \rho \tag{8}$$

It should be noted that,w_β and w_ρ are set according to [5].The estimated weights arein fact rather stable in most datasets or scenarios for their relative importance is almost fixed.

3. RESULT

In this section we show experimental results to evaluate our approach, providing comparisons with other methods on VIPeR dataset [18]. The results are shown in terms of recognition rate, by the Cumulative Matching Characteristic (CMC) curve [18]. The CMC curve represents the expectation of finding the correct match in the top n matches.Figure 2 shows the results of comparing two different patterns using spatiogram, r-spatiogram and geogram. In order to compare these methods, we use a target image in Figure 2(a) to compare its similarity with two patterns shown in Figure 2(b). We start to scan the entire image in Figure 2(b) from left to right and from top to bottom. Note that the number of subregions is set to 9, which is 3 by 3 divisions in width and height of image. The result is corresponding to any order of spatiogram and the r-spatiogram and we did not consider any assumption such as identity covariance matrix. As can be seen in Figure 2(c), peaks of similarity measure for two images are similar to each other in the result based on the spatiogram or geogram, but they are

completely distinguishable in the r-spatiogram as it is clear in Figure 2(d). In fact, we obtain same spatiograms for two images because the means and the covariances upon the coordinates of pixels are equal. Also, as the perimeter of homogenous regions are equal in the two patterns, their geograms are same. However, the r-spatiograms are not equal. It means that r-spatiogram has the discriminative capability improved in comparison with spatiogram.
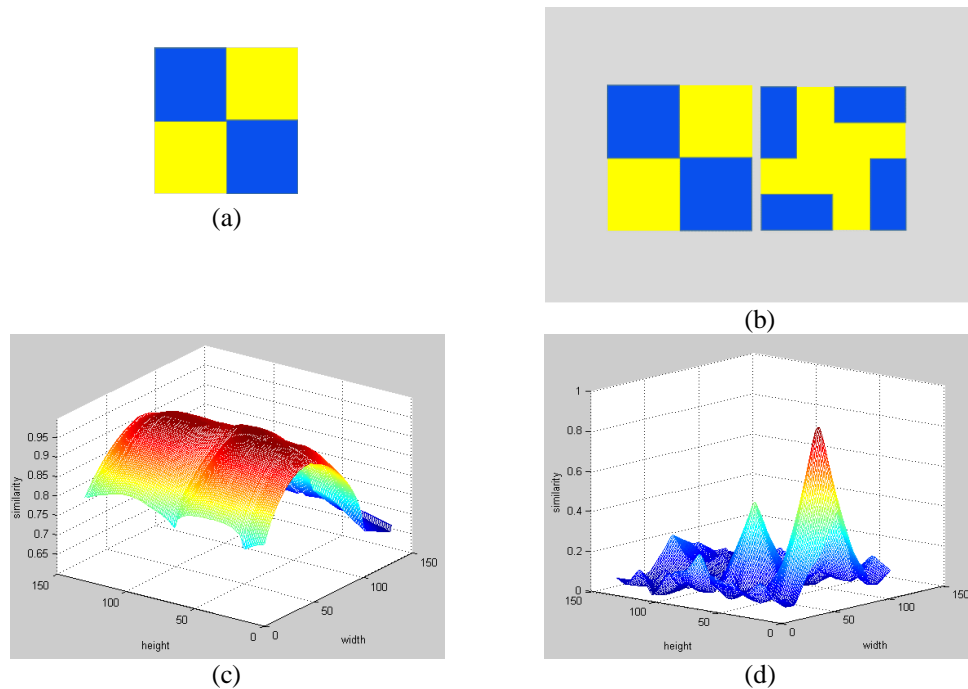

(a)

(b)

(c)

(d)

Figure 2.Results of comparing two different patterns shown in (a) and (b), using, (c) spatiogram or geogram, and (d) r-spatiogram.

Moreover, we evaluated r-spatiocovariance against other published methods, including SDALF [5], SCEAF [6], eBiCov [7], and MCTCS [7]. As can be seen in Figure 3, considering CMC curve, our descriptor outperforms other methods.
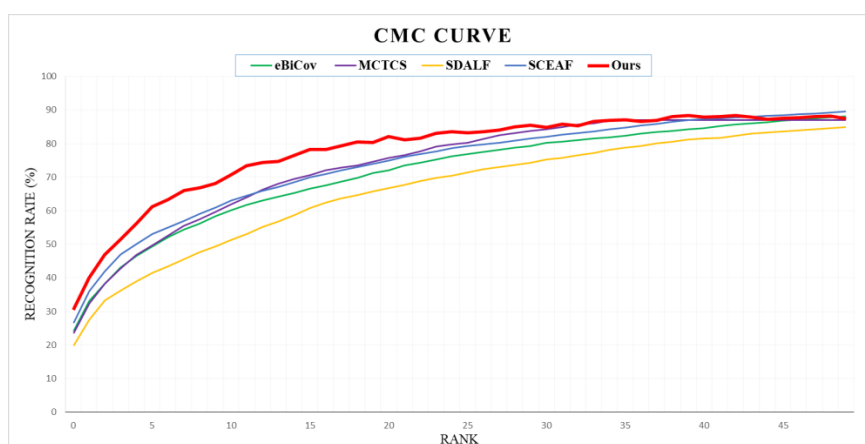

Figure 3.CMC curves on VIPeR for current methods comparison

4. CONCLUSIONS

In this paper, we have presented concept of r-spatiogram and r-spatiocovariance. Aside from their functionality which alleviate the drawbacks of traditional spatiogram and region covariance descriptors, they have constructed with using appropriate components in representing images. Therefore, such representation is more attainable and accurate. In addition, our proposed method is straightforward as it does not need any preprocessing or training steps. r-spatiogram was presented based on the region-based compactness on the distribution of the given feature and utilized in proposing our new descriptor,r-spatiocovariance which aims to provide more complete representation of image. The experimental results and performance evaluations indicated that our descriptors have the discriminative capability improved in comparison with other descriptors.

REFERENCES

[1]   Bilmes, Jeff A., (1997) "A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and hidden Markov models", *Technical Report*
.

[2]   Tuzel, Oncel, FatihPorikli, and Peter Meer, (2006) "Region covariance: A fast descriptor for detection and classification."*In European Conference on Computer Vision (ECCV),* pp589–600.

[3]   Lowe, David G., (2004) "Distinctive image features from scale-invariant keypoints",*Int. J. Comput. Vis.*, VOL. 60, NO. 2, pp91-110.

[4]   Sedaghat, Amin, and Hamid Ebadi, (2015) "Remote sensing image matching based on adaptive binning SIFT descriptor", *IEEE Trans. Geosci. Remote Sens.*, VOL. 53, NO. 10, pp5283-5293.

[5]   Farenzena, Michela, et al., (2010) "Person re-identification by symmetry-driven accumulation of local features", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).*

[6]   Hu, Yang, et al., (2013) "Exploring structural information and fusing multiple features for person re-identification", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.*

[7]   Ma, Bingpeng, Yu Su, and Frederic Jurie, (2014) "Covariance descriptor based on bio-inspired features for person re-identification and face verification"*Image and Vision Computing*, VOL. 32, NO. 6, pp379–390.

[8]   Su, Xiaoqiong, et al. (2013), "A new local-main-gradient-orientation HOG and contour differences based algorithmfor object classification", *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp2892_2895.

[9]   Birchfield, Stanley T., and SriramRangarajan,(2005) "Spatiograms versus histograms for region-based tracking", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).*

[10]   Nilsson, Mikael, et al., (2008) "On histograms and spatiograms introduction of the mapogram", *15th IEEE International Conference on Image Processing,* VOL. 3, pp973–976.

[11]   Choe, Gwangmin, et al.,(2015) "Moving object tracking based on geogram", *Multimedia Tools and Applications*, VOL.74, NO. 21, pp9771-9794.

[12]   Conaire, Ciarán O., Noel E. O'Connor, and Alan F. Smeaton,(2007), "An improved spatiogram similarity measure for robust object localisation", *International Conference on Acoustics,Speech and Signal Processing(ICASSP),* VOL. 1, ppI-1069.

[13]   Matsukawa, Tetsu, Takahiro Okabe, Einoshin Suzuki, and Yoichi Sato, (2016),"Hierarchical Gaussian descriptor for person re-identification",*IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* pp1363-1372.

[14]   Gong, Liyu, Tianjiang Wang, and Fang Liu, (2009) "Shape of Gaussians as feature descriptors", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[15]   Serra, Giuseppe, et al., (2015) "GOLD: Gaussians of local descriptors for image representation", *Computer Vision and Image Understanding*, VOL. 134 pp22–32.

[16]   Karacan, Levent, ErkutErdem, and AykutErdem,(2013) "Structure-preserving image smoothing via region covariances", *ACM Transactions on Graphics (TOG)* VOL. 32, NO. 6, 176.

[17]   Hong, Xiaopeng, et al., (2009) "Sigma set: A small second order statistical region descriptor", IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[18]   Gray, Douglas, Shane Brennan, and Hai Tao, (2007) "Evaluating appearance models for recognition, reacquisition, and tracking", *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS).*

[19]   Xu, Yuanlu, et al. (2013) "Human re-identification by matching compositional template with cluster sampling", *proceedings of the IEEE International Conference on Computer Vision (ICCV).*

Authors

Niloufar Salehi Dastjerdi received the M.Sc. degree in computer science in 2012 from University Technology Malaysia. She is currently working toward the Ph.D. degree in the department of Electrical and Computer Engineering, Concordia University, Montreal, Canada. Her research interests include visual tracking, computer vision, image processing, and pattern recognition.

M. Omair Ahmad (S'69M'78–SM'83–F'01) received the B.Eng. degree from Sir George Williams University, Montreal, QC, Canada, and the Ph.D. degree from Concordia University, Montreal, QC, Canada, both in electrical engineering. From 1978 to 1979, he was a Faculty Member with the New York University College, Buffalo, NY, USA. In September 1979, he joined the Faculty of Concordia University as an Assistant Professor of computer science. He joined the Department of Electrical and Computer Engineering, Concordia University, where he was the Chair with the department from June 2002 to May 205 and is currently a Professor. He holds the Concordia University Research Chair (Tier I) in Multimedia Signal Processing. He has authored or coauthored extensively in the area of signal processing and holds four patents. His current research interests include the areas of multidimensional filter design, speech, image and video processing, nonlinear signal processing, communication DSP, artificial neural networks, and VLSI circuits for signal processing. He was a Founding Researcher at Micronet, Ottawa, from its inception in 1990 as a Canadian Network of Centers of Excellence until its expiration in 2004. Previously, he was an Examiner of the order of Engineers of Quebec. He was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART I: FUNDAMENTAL THEORY AND APPLICATIONS from June 1999 to December 2001. He was the Local Arrangements Chairman of the 1984 IEEE International Symposium on Circuits and Systems. In 1988, he was a member of the Admission and Advancement Committee of the IEEE. He has served as the Program Co-Chair for the 1995 IEEE International Conference on Neural Networks and Signal Processing, the 2003 IEEE International Conference on Neural Networks and Signal Processing, and the 2004 IEEE International Midwest Symposium on Circuits and Systems. He was a General Co-

Chair for the 2008 IEEE International Conference on Neural Networks and Signal Processing. He is the Chair of the Montreal Chapter IEEE Circuits and Systems Society. He is a recipient of numerous honors and awards, including the Wighton Fellowship from the Sandford Fleming Foundation, an induction to Provosts Circle of Distinction for Career Achievements, and the Award of Excellence in Doctoral Supervision from the Faculty of Engineering and Computer Science of Concordia University.