FACIAL RECOGNITION, AUTONOMOUS LEARNING, BASED ON IMAGE FRAGMENTATION

Danilo Daniel González Guacte and Diego Alberto Aracena Pizarro
Department of Computer Engineering and Informatics, University Tarapacá, Chile

ABSTRACT

This paper uses variations in the image fragmentation approach, considering the number of fragments, number of samples per person, number of hyperplanes and size in the region of interest. Computational vision techniques, in particular gradient histogram (HOG) as feature space descriptor, automatic learning methods, such as vector support machines (SVM) for model construction and a face detection algorithm, are used to replicate these experiments in a real time context. Tests are performed on a population of 40 people, samples taken in uncontrolled situations, in this way some states are discarded, refining the parameters until arriving at a configuration that complies with delivering the best results in this experiment and under these conditions.

KEYWORDS

Facial Recognition, Image Fragmentation, HOG, SVM, Region of Interest, Hyperplanes

1. INTRODUCTION

At present, crime and recidivism rates reflect no less growth and, as a direct consequence, the population's distrust of its safety (Unidas, 2015) (Dammert & , 2012). This type of problem is mitigated in many ways: use of personal weapons, greater penalties for those caught, greater security personnel, among others (Dammert, 2004). One of the solutions that has given greater security to the population according to some studies (Cordero Vega, 2015) (Dammert, 2004), is use security cameras, also known as CCTV¹.

The aim is to provide a solution oriented towards use automatic recognition techniques based on biometric features. This allows the identification of individuals in a non-invasive way, since it does not need more information than that contained in the inherent traits of each individual (Viola & Jones, 2001).

The search for new ways to improve the people recognition by means of machines has been under study for several years. This work focuses on the use of fragments, motivated by the article written by: Tetyana Baydyk, Ernst Kussul and Sandra Bonilla Meza (Baydyk, et al., 2016). This paper uses SVM² for sample classification, but there are also papers that use ANN³ under the same sample fragmentation approach (Salyut & Kurnaz, 2018).

2. FACIAL RECOGNITION

Facial recognition (RF) is a technique that has been widely used even before the development of automatic recognition systems. RF has long generated the interest of the scientific community, as it has led to good levels of efficiency in this area (Hu, et al., 2017).

The face is an inherent and very discriminatory element by which individuals can be identified with the naked eye and is used for identification in various fields (Díaz Moreno, 2004). A facial recognition system can be described in five main stages (see Figure 1).

¹ Closed Circuit Television

² Support Vector Machine

³ Artificial Neural Network

Figure 1. General diagram of an object recognition system

3. SYSTEM ARCHITECTURE

The system architecture is presented in Figure 2, having as its core the processes of search and identification. These require at least two input channels (camera data and people registration).

The first input flow corresponds to the registers (learning models) of people that one tries to recognize, requires at least one sample or image per person, although there are techniques to solve this problem (Hu, et al., 2017) is not the purpose of this work. The second channel corresponds to the data provided by the camera, the objective of the application is to contrast these two inputs and through a threshold of acceptance, make a decision. If the defined threshold is exceeded, the information of the localized person is provided and in this way automate these processes.

Figure 2. General system architecture

The architecture based on layers, allows to structure the different components of the system, in such way that each component is in a layer that depends only on the lower layers and not on the upper ones. This more restrictive variant of the traditional architectural abstraction model of layers, allows the future evolution of the software through the definition of new components in higher layers. The architecture presented in Figure 3 is defined as follows:

Figure 3. Layer-based system architecture

- **Layer 1.** Level service. Layer that composes the services of the system, camera management, user management, person search, learning and notification.
- **Layer 2.** Level monitoring. That supports the monitoring component (IP camera) the component hosted in this layer requires the services of the service layer.
- **Layer 3.** Layer User. Layer that supports user interaction.

3.1 Learning Module

This module is in charge of carrying out the whole process of sample training and model generation. Figure 4 defines the learning module in 6 sub-modules: loading, detection, pre-processing, fragmentation, extraction of characteristics and classification.

Figure 4. Module diagram, for the sample training service

- **Load.** Responsible for sample reception and organization (labelling).
- **Detection.** Responsible for the detection of one face per sample (using the algorithm proposed by P. Viola y M. Jones) (Viola & Jones, 2001).
- **Pre-Processing.** In charge of "refining" the samples, he basically performs three activities: trimming, equalizing histogram and eye alignment (Viola & Jones, 2001).

- **Fragmentation.** In charge of partitioning each of the samples (four fragments, sixteen fragments or sixty-four fragments). Other works use three areas of fragmentation (eyes, nose and mouth (Baydyk, et al., 2016) (Singh, et al., 2014)).
- **Characteristics extraction.** In charge of calculating the characteristics space for each one of the fragments, using the HOG method.
- **Classification.** In charge of classifying the samples, according to the vector of characteristics and labels.

3.2 Person Search Module

Figure 5 presents the organization of the sub-modules for the search of people. This diagram shows seven sub-modules, several of which are shared with the learning module.

Figure 5. Module diagram, for the person search service

The sub-module added to this diagram is the prediction sub-module, which takes the results provided from the classification sub-module (based on the already trained models of the learning module) and evaluates the entered sample. For this, it makes use of two extra activities:

- **Merge predictions.** Accumulates predictions made by all models.
- **Prediction evaluation.** According to some criteria (75% agree in the prediction of the models (Singh, et al., 2014)) determines if this sample is possible to recognize.

4. IMPLEMENTATION

4.1 Sample Filtering

The cleaning process allows filtering those samples that will not be useful for the training of the models. The LFW⁴ database is a large facial data set containing 13.233 images, 5.749 people for an uncontrolled environment and is the data source used for static testing (Columbia, 2009) (Kumar, et al., 2009).

The following criteria are considered for this process: A single face per sample, exactly seven samples, exactly ten samples, exactly thirteen samples and more than thirteen samples. Working with 1835 samples for test execution.

⁴ Labeled Faces in the Wild

4.2 Fragmentation of Samples

The fragmentation process is worked with at least three degrees of fragmentation, factor 4, 16 and 64. This means that for factor 4, a face is fragmented into four parts and for each of these fragments, a learning machine is trained, specialized in the region of interest assigned (see Figure 6).

Figure 6. Example of face fragmentation

4.3 Calculation of HOG Descriptors

To use the HOG method as descriptor, enter a grayscale image (matrix of pixels with values between 0-255). The output of this method delivers what is known as vector descriptor or feature vector. Each of the components of this vector reflects the final value of the histogram associated with a cell (equation 1). The dimension of the vector depends on the number of blocks, number of cells and the orientation angle, with or without sign.

$$h((k) = \sum_{(x,y) \in C} w_k(x, y) * g(x, y) \quad (1)$$

This means that the calculation of the histogram is given by the sum of each gradient multiplied by a weighting factor associated with the pixel. Tables 1, 2 and 3 present the configuration used for three region sizes of interest tested, depending on each degree of fragmentation.

Table 1. Parameters for a HOG object, with degree of fragmentation 4

Roi (px)	Win size (px)	Block size (px)	Cell size (px)	N° bins
64x64	32x32	16x16	8x8	9
128x128	64x64	32x32	16x16	9
64x128	32x64	16x32	8x16	9

Table 2. Parameters for a HOG object, with degree of fragmentation 16

Roi (px)	Win size (px)	Block size (px)	Cell size (px)	N° bins
64x64	16x16	8x8	4x4	9
128x128	32x32	16x16	8x8	9
64x128	16x32	8x16	4x8	9

Table 3. Parameters for a HOG object, with degree of fragmentation 64

Roi (px)	Win size (px)	Block size (px)	Cell size (px)	N° bins
64x64	8x8	4x4	2x2	9
128x128	16x16	8x8	4x4	9
64x128	8x16	4x8	2x4	9

5. RESULTS

Tests are performed on a random population of 40 people, the categories are: seven samples per person, ten samples per person, thirteen samples per person and more than thirteen samples per person (and less than 20 samples). According to this categorization, tables 4, 5, 6 and 7 describe the recognition success rate, considering only the number of hyperplanes and degrees of fragmentation.

Table 4. Average of results carried out on the population, for category 7 samples per person, evaluating degrees of fragmentation and number of hyperplanes

N° hyperplanes	Frag. 4 (%)	Frag. 16 (%)	Frag. 64 (%)	Score
4	60	80	80	220
9	65	65	50	180
14	43,33	50	40	133,33
19	40	40	40	120

Table 5. Average of results carried out on the population, for category 10 samples per person, evaluating degrees of fragmentation and number of hyperplanes

N° hyperplanes	Frag. 4 (%)	Frag. 16 (%)	Frag. 64 (%)	Score
4	45	60	60	165
9	60	70	45	175
14	40	63,33	36,67	140
19	47,5	55	32,5	135

Table 6. Average of results carried out on the population, for category 13 samples per person, evaluating degrees of fragmentation and number of hyperplanes

N° hyperplanes	Frag. 4 (%)	Frag. 16 (%)	Frag. 64 (%)	Score
4	55	55	35	145
9	60	60	45	165
14	56,67	50	43,33	150
19	45	45	30	120

Table 7. Average of results carried out on the population, for category +13 samples per person, evaluating degrees of fragmentation and number of hyperplanes

N° hyperplanes	Frag. 4 (%)	Frag. 16 (%)	Frag. 64 (%)	Score
4	80	100	60	240
9	65	70	45	180
14	56,67	66,67	40	163,33
19	52,5	57,5	40	150

These tests suggest the use of between 4 to 9 hyperplanes per learning machine, i.e. 4 to 10 people per model. Tables 8, 9 and 10 describe the success rate considering the size of the region of interest (64x64 pixels, 64x128 pixels and 128x128 pixels) and the degree of fragmentation.

Table 8. Average results obtained for tests carried out on a degree of fragmentation 4

Roi (px)	7 samples (%)	10 samples (%)	13 samples (%)	+13 samples (%)	Score
64x64	65	60	60	65	250
128x128	75	60	60	70	265
64x128	60	55	50	70	235

Table 9. Average results obtained for tests carried out on a degree of fragmentation 16

Roi (px)	7 samples (%)	10 samples (%)	13 samples (%)	+13 samples (%)	Score
64x64	65	70	60	70	265
128x128	60	55	65	70	250
64x128	70	55	55	70	250

Table 10. Average results obtained for tests carried out on a degree of fragmentation 64

Roi (px)	7 samples (%)	10 samples (%)	13 samples (%)	+13 samples (%)	Score
64x64	50	45	45	45	185
128x128	55	50	55	50	210
64x128	55	45	65	65	230

These results make it possible to approximate a suitable type of configuration when working with HOG descriptors and SVM-based learning machines. Using an orientation based on fragments, number of samples per person, number of hyperplanes per learning machine and size of the region of interest. In this sense and supported by the different tests carried out, it is possible to conclude a suitable configuration for the parameters that were part of the experimentation (see Table 11).

Table 11. Result of better configuration

Parameter	Configuration
Fragmentation	16
N° hyperplanes	4-9
N° Samples	+13
Region of interest	64x64

6. CONCLUSIONS

This work allows to know an effective configuration, verified with samples of people in an uncontrolled context. Allowing to establish the number of fragments, number of samples per person and size of the region of interest that provides the best results for people recognition.

The use of HOG as a method to describe an image is mainly due to the great use of it in different approaches in recognition work, its low computational cost and because it was necessary to opt for some method and lay the groundwork to empirically demonstrate the use of image fragments. Being a method that works on contours in an image, it was possible to predict that, by fragmenting one more face, the HOG method would not have many features to highlight and, therefore, the models would not be robust.

Likewise, the parameters for the configuration of each learning machine were adjusted thinking about a later application in real time, so a best option is to use a linear kernel to find an optimal hyperplane and with certain slack or slack variables. While this configuration is not the best for sample separation, it has a low computational cost. From this point of view, the integration of more CPUs or even GPUs could be worked on, since the fragmentation process allows exclusive work between fragments (parallelization of processes). This would allow to work with a better image resolution, and use methods that require a higher computing power and not be so restricted by the hardware.

REFERENCES

- Baydyk, T., Kussul, E. & Meza, S. B., 2016. Facial Recognition on the Basis of Facial Fragments. *International Journal of Computer and Information Engineering*, 10(12).
- Columbia, U., 2009. www.cs.columbia.edu. [Online] Available at: <http://www.cs.columbia.edu/CAVE/databases/pubfig/> [Accessed 13 05 2019].
- Cordero Vega, L., 2015. Videovigilancia e intervención administrativa: las cuestiones de legitimidad. *revistatrabajo.uchile*, pp. 359-376.
- Dammert, L., 2004. ¿Ciudad sin ciudadanos? Fragmentación, segregación y temor en Santiago. *EURE*, 30(91), pp. 87-96.
- Dammert, L. & L. M., 2012. *La Seguridad Ciudadana: El problema principal de América Latina*. Santiago de Chile: Corporación Latinbarómetro.
- Díaz Moreno, A. B., 2004. *Reconocimiento Facial Automático Mediante Técnicas de Visión Tridimensional*. Madrid: Universidad Politécnica de Madrid.
- Hu, X., Peng, S., Wang, L. & Lo, Z., 2017. Surveillance video face recognition with single sample per person based on 3D modeling and blurring. *Elsevier*, pp. 113-117.

- Kumar, N., Berg, A. C., Belhumeur, P. N. & Nayar, S. K., 2009. Attribute and Simile Classifiers for Face Verification. *International Conference on Computer Vision (ICCV)*, pp. 365-372.
- Salyut, J. & Kurnaz, C., 2018. Profile Face Recognition using Local Binary Patterns with Artificial Neural Network. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pp. 1-4.
- Singh, A. K., Kumar, A., Nandi, G. C. & Chakroborty, P., 2014. Expression invariant fragmented face recognition. *2014 International Conference on Signal Propagation and Computer Technology*, pp. 184-189.
- Singh, A. K., Kumar, A., Nandi, G. C. & Chakroborty, P., 2014. Expression invariant fragmented face recognition. *2014 International Conference on Signal Propagation and Computer Technology*, pp. 184-189.
- Unidas, N., 2015. *Naciones Unidas*. [Online] Available at: <http://www.un.org/es/sections/issues-depth/population/index.html> [Accessed 13 05 2019].
- Viola, P. & Jones, M., 2001. Robust Real-Time Face Detection. *Computer Vision and Pattern Recognition*, pp. 747-747.