Head Pose Classification Using a Bidimensional Correlation Filter

Djemel Ziou[1(✉)], Dayron Rizo-Rodriguez[1], Antoine Tabbone[2], and Nafaa Nacereddine[3]

[1] Départment d'Informatique, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada
djemel.ziou@usherbrooke.ca
[2] LORIA UMR 7503, Université de Lorraine, BP 239, 54506 Vandoeuvre-les-nancy, France
[3] Centre de Recherche Scientifique et Technique en Soudage et Controle Chéraga, 16002 Algiers, Algeria

Abstract. Correlation filters have been extensively used in face recognition but surprisingly underused in head pose classification. In this paper, we present a correlation filter that ensures the tradeoff between three criteria: peak distinctiveness, discrimination power and noise robustness. Such a filter is derived through a variational formulation of these three criteria. The closed form obtained intrinsically considers multiclass information and preserves the bidimensional structure of the image. The filter proposed is combined with a face image descriptor in order to deal with pose classification problem. It is shown that our approach improves pose classification accuracy, especially for non-frontal poses, when compared with other methods.

Keywords: Head pose classification · Bidimensional correlation filter

1 Introduction

The orientation of a subject's head relative to a given coordinate systems is commonly referred as head pose in computer vision [11]. The pose of a human head in such a system depends on the muscular movement of the neck as well as the orientation of the body. Pose estimation can be useful in multiple practical applications including face recognition, gaze direction estimation and hands-free human computer interaction. Taking into account the level of precision, we can carry out either a fine or a coarse pose estimation. In the first case, we aim at determining the pitch, roll and yaw rotation angles by assuming that human head is limited to three degrees of freedom [11]. In the second case, a finite number of poses are modeled so that pose estimation becomes a classification problem.

Pose estimation problem from both perspectives has been tackled using different types of approaches. The reader can find further details about the taxonomy

M. Kamel and A. Campilho (Eds.): ICIAR 2015, LNCS 9164, pp. 203–209, 2015.
DOI: 10.1007/978-3-319-20801-5_22

of these methods in [11]. In this work, we address in particular the appearance-based category for coarse pose classification [1,2,9,16–18]. Appearance-based approaches extract facial features assuming that such features are related to the pose of the head [9]. As a result, a statistical learning technique is used from the features extracted in order to infer each pose class. According to [11], these methods are advantageous because they do not require facial points to be detected. However, their most significant drawback is the fact that a different identity can produce more dissimilarity than a change in pose classification.

An interesting technique that has not been widely exploited for pose classification are correlation filters, especially considering that they are quite popular in face recognition [4,7,8]. A correlation filter represents a template for recognizing a specific pattern. Thus, a sharp peak is expected to indicate the presence of the such a pattern when correlating the filter with a target image. Such a technique allow us to operate in frequency domain, beneficial for recognition problems, while providing shift-invariance, noise robustness and a closed form solution [7]. Consequently, a correlation filter can be advantageous for coping with the problem of appearance-based pose classification.

A well-know filter is the Optimal Trade-off Filter (OTF) [13], which is derived from the constrained optimization of two criteria: Signal-to-Noise Ratio (SNR) and the Peak-to-Correlation Energy (PCE). Other filters (MVSDF [6], ECPSDF [12], MACE [10], and POUMACE [8]) are special cases of the OTF. However, most of these correlation filters are designed by transforming the image into a one-dimensional vector. It means that the spatial structure of the image is somehow neglected. In this work, a correlation filter is derived from three criteria through a variational formulation. As a result, we obtain a close form preserving image spatial structure unlike existing correlation filters. The correlation filter proposed is compared with state-of-the-art approaches for face pose classification. The rest of the paper is structured as follows. In Sect. 2, the derivation of the filter is presented. Experimental results are shown in Sect. 3. Finally, the conclusions are drawn in Sect. 4.

2 Derivation of 2D Correlation Filter

In this section, we propose a new correlation filter $h_k(X)$ where $X = (x, y)$ represents space domain coordinates. For this purpose, we assume that common features are mostly shared within the same class k and not with other classes. Firstly, we define the three criteria to be fulfilled by our correlation filter. Secondly, a closed form is obtained by using variational calculus.

The first criteria represents that a correlation filter should deliver a sharp peak at the origin for an image belonging to the correct class. For this reason, we force the correlation space to resemble a known function $g_k(X)$ having the maximal energy in the origin such as: Gaussian function or a Dirac distribution. It would be then defined as follows:

$$min_{h_k} \oint_{-\infty}^{+\infty} ((v_{ik} \copyright h_k)(X) - g_k(X))^2 dX \tag{1}$$

where the correlation space $(v©h)(X)$ represents an inner product of the image $v(X)$ and the shifted correlation filter $h(X)$.

The second criteria imposes that the correlation energy between a filter and an image of the same class should be maximal while being minimal otherwise. Then, it can be simply expressed as the minimization of the correlation energy with respect to other classes:

$$min_{h_k} \sum_{l \neq k}^{M} \sum_{j=1}^{N_l} \oint_{-\infty}^{+\infty} (v_{jl} © h_k)^2(X) dX \tag{2}$$

The third requirement focuses on the effects of noise in filter performance. According to Rice [14], noise produces several maxima in the correlation surface. Setting the distance between maxima to a specific constant reduces the number of maxima in the correlation space. It has been shown in [3], that such a requirement is equivalent to derive a filter that ensures a smooth correlation space. Then, the smoothness can be measured through the norm of the field produced by a differentiation operator applied to the correlation space:

$$min_{h_k} \oint_{-\infty}^{+\infty} ||\nabla(v_{ik} © h_k)(X)||^2 dX \tag{3}$$

Therefore, the variational problem can be formulated as the minimization of the weighted sum of the three criteria. Taking into account all images of the k^{th} class, it can be written as:

$$\begin{aligned} h_k = argmin_{h_k} \oint_{-\infty}^{+\infty} & (\sum_{i=1}^{N_k} ((v_{ik} © h_k)(X) - g_k(X))^2 \\ & + \lambda_2 \sum_{i=1}^{N_k} ((v_{ik} © h_{xk})(X)^2 + (v_{ik} © h_{yk})(X)^2) \\ & + \lambda_1 (\sum_{l \neq k}^{M} \sum_{j=1}^{N_l} (v_{jl} © h_k)^2(X))) dX \end{aligned} \tag{4}$$

where λ_1 and λ_2 are Lagrangian multipliers, N_k is the number of samples for the k^{th} class, and $h_{kx}(X)$ and $h_{ky}(X)$ are the first order derivatives of $h_k(X)$. In this variational formulation, we notice the multiclass nature of this correlation filter, i.e., it depends on both the images of the k^{th} class and those of the other classes. Moreover, the bidimensional structure of the images is preserved because they are never transformed into a column vector. This is a distinctive point of our proposal when compared to traditional correlation filters.

In order to derive the closed form of the filter $h_k(X)$ solving the variational problem, Gateaux derivatives are used to find the Euler-Lagrange partial differential equation (PDE). A straightforward manipulation of the PDE in Fourier domain results in the following closed form (see Appendix):

$$\hat{h}_k(W) = \frac{\hat{g}_k(W)}{\sum_{i=1}^{N_k} ||\hat{v}_{ik}(W)||^2 (1 + \lambda_2 ||W||^2) + \lambda_1 \sum_{l \neq k}^{M} \sum_{j=1}^{N_l} ||\hat{v}_{jl}(W)||^2} \sum_{i=1}^{N_k} \hat{v}_{ik}^*(W)$$

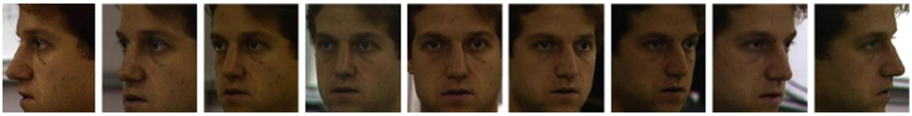

Fig. 1. Pose classes in CMU PIE database (Left:-90° to Right: 90°)

where $\hat{}$ denotes the Fourier Transform, * the complex conjugate and $W = (w_x, w_y)$ represents the frequency domain coordinates.

3 Experimental Results

In the experiments, the CMU PIE database [15] is employed for comparing the correlation filter proposed with state-of-the-art approaches in pose classification. This database was constructed from 68 individuals, each one captured under 9 different horizontal poses (yaw) and 21 illumination conditions. Since we are evaluating pose variations, only the 612 images frontally illuminated will be used in our experiments. Those poses range from -90° to 90° using a 22.5° step, see Fig. 1. From all these images, the faces were automatically detected and cropped to a 32 x 32 resolution based on the position of the eyes. Then, a half of the images per each pose class are randomly selected for training, 306 in total, and the other 306 are used for testing. This two-fold cross-validation is the same experimental setup used for testing the other three methods to which we will compare our approach. Additionally, we use the Local Binary Patterns (LBP) descriptor of the face images for designing the corresponding correlation filter because LBP has been effective in face recognition [5]. It should be noticed that we refer to the LBP descriptor as an image, i.e. we do not compute the histogram, so that the bidimensional structure is preserved. This idea of using a face descriptor instead of the raw image is also exploited in [2] but employing Gabor features.

Table 1. Pose estimation accuracy (%) per Class for 32 x 32 resolution.

		Per pose class								
	Mean	**-90°**	**-67.5°**	**-45°**	**-22.5°**	**0°**	**22.5°**	**45°**	**67.5°**	**90°**
Brown(Probabilistic) [2]	91	–	–	–	–	–	–	–	–	–
Brown(Neural Network) [2]	91	–	–	–	–	–	–	–	–	–
Takallou [16]	90.1	85	87	91	87	94	97	89	**93**	88
LBP+2DCorrFilter	**94.38**	**91.18**	**91.18**	**95.88**	**95.29**	**95.88**	**100**	**94.71**	89.41	**95.88**

In Table 1, we present the performance delivered by pose estimation approaches in CMU PIE database. The first column shows the mean score, i.e. considering all pose classes, and the other columns the specific score per class if provided. In the case of the state-of-the-art methods, we simply display the percentages reported by each of them. For this reason, we limit our comparison to those pose classification approaches that has been tested in CMU PIE database following the experimental setup described above. For our proposal, the average pose estimation accuracy computed from 10 different two-fold cross-validation configurations is reported.

On one hand, it can be seen in the first column that the LBP+2DCorrFilter approach improves the other methods in more than 3 % in terms of the mean score. The best performance of the 2D correlation filter is obtained by setting $\lambda_1 = \lambda_2 = 0.5$ and using a Gaussian function as $g_k(X)$. On the other hand, we appreciate that our method is superior to Takallou's approach in 8 out of the 9 pose classes considered. For frontal pose, $\mathbf{0}^\circ$, both of them deliver comparable percentages. In the case of $\mathbf{67.5}^\circ$ pose, Takallou's method overcomes LBP+2DCorrFilter in about 4 %. However, in the remaining pose classes the difference in favour of LBP+2DCorrFilter combination is clearly observed reaching a maximal gap of about 8 % for the $\mathbf{90}^\circ$ pose class.

Overall, we can see that the LBP+2DCorrFilter method delivers the top mean pose classification score among the approaches compared. More importantly, our approach is capable of overcoming Takallou's approach for most of the pose classes deviated from the frontal pose. It means that preserving bidimensional image structure actually contributes to improve pose estimation when dealing with non-frontal head orientations. In terms of computational complexity, the design of the correlation filters representing each class is $O(Nd(1+\log d))$ where N the total number of training images and d the dimension of one image. It means that the complexity of our proposal is based on the computation of N Fourier Transforms which can be executed in $O(d \log d)$. If we consider that the correlation filters can be derived off-line, such a cost is reduced to computing one Fourier Transform for the target image in recognition phase.

4 Conclusions

In this work, a bidimensional correlation filter is presented. It is derived from three criteria which take into account the peak at the origin, the discrimination power and the robustness to noise of the filter. The closed form obtained through variational calculus has a multiclass nature and preserves the bidimensional structure of an image representation. In the experimental evaluation, we combine the effective LBP descriptor with the correlation filter proposed for head pose classification. When compared with other state-of-the-art methods, our approach delivers the top performance in terms of the mean percentage as well as for most of the nine pose classes evaluated.

5 Appendix

Let us now compute Gateaux derivative of the variational problem in Eq. (4).

$$\frac{\partial}{\partial \varepsilon} E^k(h_k + \varepsilon \varphi_k))|_{\varepsilon=0} = \oint_{-\infty}^{\infty} (\sum_{i=1}^{N_k} ((v_{ik} \copyright h_k)(X) - g_k(X))(v_{ik} \copyright \varphi_k(X))$$

$$+\lambda_2 \sum_{i=1}^{N_k} ((v_{ik} \copyright h_{xk})(X)(v_{ik} \copyright \varphi_{xk})(X) + (v_{ik} \copyright h_{yk})(X)(v_{ik} \copyright \varphi_{yk})(X))$$

$$+\lambda_1 \sum_{l \neq k}^{M} 2 \sum_{j=1}^{N_l} (v_{jl} © h_k)(X)(v_{jl} © \varphi_k)(X))dX = 0$$

The first term of this equation can be rewritten as:

$$\oint_{-\infty}^{\infty} \sum_{i=1}^{N_k} ((v_{ik} © h_k)(X) - g_k(X))(v_{ik} © \varphi_k)(X)dX$$
$$= \oint_{-\infty}^{\infty} \varphi_k(t) \sum_{i=1}^{N_k} (v_{ik} ©((v_{ik} © h_k) - g_k)(t)dt$$

We apply the same analysis for second and third terms. Then, the result of the second term is integrated by parts. As a result, we get the following Euler-Lagrange equation:

$$\sum_{i=1}^{N_k} (v_{ik} © v_{ik} © h_k)(X) + \lambda_2 \sum_{i=1}^{N_k} ((v_{ik} © v_{ik} © h_{xk})_x(X) + (v_{ik} © v_{ik} © h_{yk})_y(X))$$
$$+ \lambda_1 \sum_{l \neq k}^{M} \sum_{j=1}^{N_l} (v_{jl} © v_{jl} © h_k)(X) = \sum_{i=1}^{N_k} (v_{ik} © g_k)(X)$$

A straightforward manipulation allows to write:

$$\sum_{i=1}^{N_k} (v_{ik} © v_{ik} © h_k)(X) + \lambda_2 \sum_{i=1}^{N_k} ((v_{xik} © v_{xik} © h_k)(X) + (v_{yik} © v_{yik} © h_k)(X))$$
$$+ \lambda_1 \sum_{l \neq k}^{M} \sum_{j=1}^{N_l} (v_{jl} © v_{jl} © h_k)(X) = \sum_{i=1}^{N_k} (v_{ik} © g_k)(X)$$

This equation is solved in the Fourier domain by using the properties $\hat{a}\hat{a}^* = ||\hat{a}||^2$ and $f_x\hat{(X)}(W) = jw_x\hat{f}(W)$, the resulting filter is given by:

$$\hat{h}_k(W) = \frac{\sum_{i=1}^{N_k} \hat{v}_{ik}^*(W)\hat{g_k}(W)}{\sum_{i=1}^{N_k} ||\hat{v}_{ik}(W)||^2(1 + \lambda_2||W||^2) + \lambda_1 \sum_{l \neq k}^{M} \sum_{j=1}^{N_l} ||\hat{v}_{jl}(W)||^2}$$

where * denotes the complex conjugate.

References

1. Ba, S.O., Odobez, J.M.: A probabilistic framework for joint head tracking and pose estimation. In: Proceedings of ICPR 2004. vol. 4, pp. 264–267. IEEE (2004)
2. Brown, L.M., Tian, Y.L.: Comparative study of coarse head pose estimation. In: Proceedings Workshop on Motion and Video Computing, 2002. pp. 125–130. IEEE (2002)

3. Canny, J.: A computational approach to edge detection. IEEE TPAMI PAMI **8**(6), 679–698 (1986)
4. Heo, J., Savvides, M., Vijayakumar, B.V.K.: Advanced correlation filters for face recognition using low-resolution visual and thermal imagery. In: Kamel, M.S., Campilho, A.C. (eds.) ICIAR 2005. LNCS, vol. 3656, pp. 1089–1097. Springer, Heidelberg (2005)
5. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local binary patterns and its application to facial image analysis: a survey. IEEE Trans. Syst. Man Cybern. **41**(6), 765–781 (2011)
6. Kumar, B.V.K.V.: Minimum-variance synthetic discriminant functions. J. Opt. Soc. Am. A **3**(10), 1579–1584 (1986)
7. Kumar, B.V., Savvides, M., Xie, C.: Correlation pattern recognition for face recognition. Proc. IEEE **94**(11), 1963–1976 (2006)
8. Levine, M.D., Yu, Y.: Face recognition subject to variations in facial expression, illumination and pose using correlation filters. CVIU **104**(1), 1–15 (2006)
9. Ma, B., Chai, X., Wang, T.: A novel feature descriptor based on biologically inspired feature for head pose estimation. Neurocomputing **115**, 1–10 (2013)
10. Mahalanobis, A., Kumar, B.V.K.V., Casasent, D.: Minimum average correlation energy filters. Appl. Opt. **26**(17), 3633–3640 (1987)
11. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–626 (2009)
12. Ng, C.: PDA Face Recognition System Using Advanced Correlation Filters. Master thesis, Carnegie Mellon University (2005)
13. Refregier, P.: Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and horner efficiency. Opt. Lett. **16**(11), 829–831 (1991)
14. Rice, S.O.: Mathematical analysis of random noise. Bell Syst. Tech. J. **23**, 282–332 (1944)
15. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression (pie) database. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–51 May 2002
16. Takallou, H.M., Kasaei, S.: Head pose estimation and face recognition using a non-linear tensor-based model. IET Comput. Vision **8**(1), 54–65 (2014)
17. Tian, Y., Brown, L., Connell, J., Pankanti, S., Hampapur, A., Senior, A., Bolle, R.: Absolute head pose estimation from overhead wide-angle cameras. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG 2003, pp. 92–99. IEEE (2003)
18. Yan, D., Yan, Y., Wang, H.: Robust head pose estimation with a new principal optimal tradeoff filter. In: Sun, C., Fang, F., Zhou, Z.-H., Yang, W., Liu, Z.-Y. (eds.) IScIDE 2013. LNCS, vol. 8261, pp. 320–327. Springer, Heidelberg (2013)