Relative Spatial Weighting of Features for Localizing Parts of Faces

Jacopo Bellati and Díbio Leandro Borges

Department of Computer Science, University of Brasilia,
70910-900 Brasília, DF, Brazil
bellati@gmail.com, dibio@unb.br

Abstract. This paper proposes an approach for detecting important parts of faces in uncontrolled imaging settings. Regions of special interest in faces of humans are eyes and eyebrows, nose and mouth. The approach works by first extracting ORB (Oriented FAST and Rotated BRIEF) and SURF (Speeded up robust features) features, secondly a supervised learning step with a random subset of images is performed using k-means algorithm for devising the clusters' centers of the important parts of faces. For the testing set of images the normalized values of each new ORB or SURF feature is weighted positively depending on its similarity and proximity of a cluster center (a face part). Tests were performed using the BioID dataset which consists of 1521 images of 23 different subjects in a variety of situations. Results show that the use of ORB features for face parts localization is more efficient and more precise than SIFT or SURF features alone. Also, the relative spatial weighting of a combination of ORB and SURF features enhances the localization of parts of faces.

Keywords: Face parts localization, ORB features, face detection.

1 Introduction

Face detection and related applications have been at the top of approached problems by the Computer Vision and Pattern Recognition research community. A step forward to be taken is to have more detailed localization of face parts, or facial features as they are also called, since identity, sorting and editing face applications are dependent on dealing with the face parts in separate. 2D face parts can be considered as facial feature points, as specifying for example center of eye, tip of nose, mouth corners, as in [2] [6] [5], or as large scale 2D facial parts such as eyes and eyebrows, nose and mouth, as in [7] [9]. In this paper we approach the 2D face parts localization problem considering four large scale facial parts: eye and eyebrow right, eye and eyebrow left, nose, and mouth. Regarding a recent taxonomy of facial features proposed in [11] our work explores level 2 features, which are locally derived and can describe structures relevant to face recognition.

Most representative of face fiducial points works are [2], [6], and [5]. In [2] they formulate the problem of part localization as Bayesian inference combining local detectors and a prior model of face shape. They use a large collection of exemplars, and

the part locations are decided on a consensus (RANSAC based) decision to disambiguate candidates. They present tests on the BioID database and a proprietary one. [6] presented a method to detect face fiducial points based on regression forests. 13000 images annotated with 10 fiducial points are trained and the ensembles of regression trees estimate the positions of the fiducial points. [5] presented a facial landmark (fiducial points) localization method based on Haar features and gradient boosted trees to predict the landmark positions, 9 landmark positions are previously defined and results are shown on subsets of BioID and their own image database.

More related to the approach proposed here are [9] and [7]. [9] investigated detection of large scale facial features by using an appearance based feature vector of gaussian derivatives of normalized face images. The facial features are defined as salient from the face images. Results were presented with 30 images detected eyes, nose, mouth and chin as important facial features, and were very dependent on scale, and invariance was not considered. [7] proposes to detail and detect facial features (eyes, nose, eyebrows, mouth, chin) by constructing appearance vectors of the features, and also of the context surrounding the features. A supervised learning discriminative algorithm is then applied to classify features and non-features samples. Results were shown for 1200 face images with uniform and controlled background with error rates below 5%. Detection of facial features on varying lighting and background conditions were not shown.

In this paper we propose a relative spatial weighting algorithm for localization of face parts. One of our motivations is to explore feature detector such as ORB [15] and SURF for face parts localization, and to investigate facial feature detection for face identification applications. Our main contributions in this work are: 1) provide a learning algorithm to select ORB and SURF features for face parts localization; 2) results on a benchmark image database for face parts localization.

2   Invariant Feature Descriptors

Finding correspondences between different images of the same object, considering a variety of lighting, viewing and scaling conditions is a major task in Computer Vision and applications.   A large variety of feature descriptors and their respective matching algorithms have been proposed in the literature. One of the most successful is the SIFT descriptor (Scale Invariant Feature Transform)[13], although more recently promising invariant feature descriptors have been presented such as SURF (Speeded Up Robust Features)[1], and ORB (Oriented FAST and Rotated BRIEF) [15].

2.1   SIFT Descriptor

SIFT is a highly distinctive, scale and rotation invariant descriptor. It is computed in four main steps [13]: 1) Extract the keypoints from the image as local extrema (minima or maxima) using a Difference of Gaussians (DoG) Pyramid. Each point is compared to its 8 neighbors in the same scale, as well as to its 9 neighbors in the upper and lower scale; 2) Localize the keypoints, position and scale, by fitting a quadratic polynomial and rejecting weak keypoints by a Hessian matrix curvature test. 3) Surrounding a

keypoint compute a histogram of gradient directions , and then assign the canonical orientation (single or multiple) of the patch as the peak(s) of the smoothed histogram; 4) The keypoint descriptor is formed by a 128 vector of 16 histograms with 8 orientations, considered in 16x16 windows with keypoint at center. Matching can be performed by comparing two descriptors with a distance function.

2.2   SURF Descriptor

SURF has been devised to be a faster and more robust descriptor, and matcher feature algorithm, than SIFT [1].   First it approximates the derivatives of Hessian matrix by box filters and uses the integral image as basis for computations. The determinant of H is also used for keypoint localization which is weighted to obtain a good approximation. Orientation assignment is done by evaluating a circular neighborhood around the keypoint and computing haar horizontal and vertical responses using the integral image also as basis.   The SURF descriptor considers square regions and sum the responses (vertical and horizontal) for each subregion separately. A vector of 128 elements of those sums for keypoint regions is formed as the descriptor. It has been reported [15] that SURF is one order of magnitude faster than SIFT, however less robust to viewpoint and illumination changes.

2.3   ORB Descriptor

ORB descriptor [15], Oriented FAST and Rotated BRIEF, builds on good properties of FAST and BRIEF descriptors.   Two main innovations are made on them, first it adds to a FAST descriptor an orientation computation by a weighted averaging of pixel intensities in the local patch. This centroid operator gives a single dominant orientation. Second it uses an ID3 machine learning algorithm for de-correlating BRIEF features under rotational invariance, and this is used for sampling point pairs to the descriptor. ORB is a binary descriptor, aimed to be an efficient alternative to SIFT or SURF descriptors [15].   Matching can be computed by Hamming distance. ORB has been reported [15] to be about 10 times faster than SURF, 100 times faster than SIFT, and less sensitive to gaussian noise than SIFT.

3   Face Parts Description and Localization

Face detection and face recognition are tasks with great interest from the research community. One of the most important subtasks of it is the identification and localization of important face parts, or facial features, as eyes, eyebrows, nose and mouth. Many applications besides recognition of individuals, such as autofocus, white balancing, sorting and retrieving face images, semi-automatic editing, depend on the localization of the face parts.

There are two basic different approaches for the localization of face parts: one that considers the facial feature points as relevant elements to be identified [2] [5] [6], a variation from 5 to 30 points have been reported in the literature as those facial feature

points, for example eyes and mouth corners, centers, and middle points of contours and nose; and another that considers large-scale facial parts or the whole region of interest of eyes, nose and mouth as a facial part [7] [9]. The approach considered in this paper is the second one, where the face parts are four main regions: left eye and respective eyebrow, right eye and respective eyebrow, nose, and mouth. Figure.1 shows an image from the BioID database used as a training image with these four facial parts marked on it.
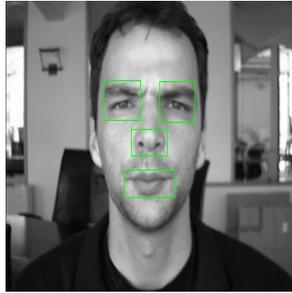

Fig. 1. One training image from the BioID database with the four proposed face parts shown

State of the art feature detectors such as SIFT [13], SURF [1], and more recently ORB [15] have been applied and demonstrated impressive performance on challenging recognition and tracking tasks [14]. However, only in the case of SIFT [12] [3], and SURF [8] there had been some attempts to address face identification and recognition using it as descriptor. To the best of our knowledge ORB has not been applied on the mentioned problems here. It is a hypothesis of this work that face parts localization can be addressed by one, or a combination of these feature detectors, especially the ORB since it shows top properties [15] a descriptor is aimed to demonstrate. Figure.2 shows a typical frontal face, from a benchmark image database for face identification BioID [4], with marked the 50 most salient feature points output by (a) ORB, (b) SIFT, and (c) SURF detectors.

Since one aim is to localize face parts from typical images considering illumination variations, relevant backgrounds, and face variations by expressions such as talking, smiling, closing eyes, wearing glasses, it can be shown that these three feature detectors have different responses from each other.

This work proposes to learn from a set of images, a supervised training set of the closest features (ORB and SURF) to the face parts, and then classify other images for face parts by devising a k-means with relative weighting for the trained centroids of the parts. Initial tests had shown that SIFT, besides being the much slower detector, picked the farthest from the face parts aimed. It is a hypothesis from this work that ORB and SURF features can be used for an efficient face parts localization. A novel algorithm for performing such selection and localization is presented next.

<div align="center">(a)   (b)   (c)</div>
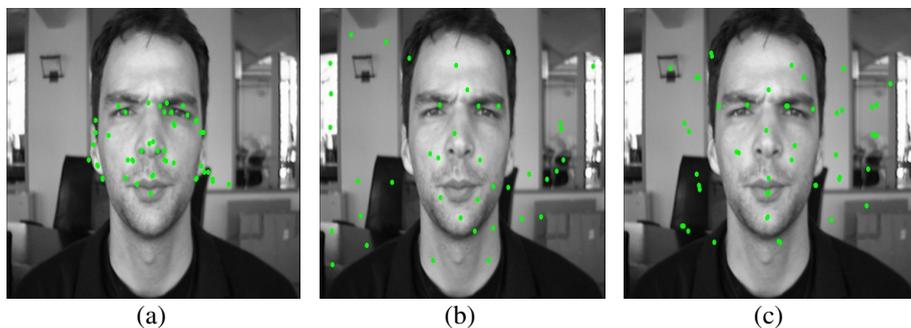
Fig. 2. Example images from the BioID database with (a) 50 highest ORB features shown; (b) 50 highest SIFT features shown; (c) 50 highest SURF features shown

4   Relative Spatial Weighting of Features

We propose to train and select a subset of features (ORB and SURF) in order to have the large scale face parts being localized and sampled for sure. For this we devise the following algorithm which positively weights the closest features to the facial parts, and negatively the farthest. First, it selects a subset of images for training, label the most salient normalized features (up to 30 ORB and SURF) belonging to the face parts; Average those feature outputs to each face part and keep its statistics (mean values and relative distances of the centroids); For a new image, compute and normalize the 30 most salient features (ORB and SURF); then for each feature, from the most to the least salient, run a k-means having as seeds the trained face parts; if a feature is decided close and similar enough (e.g. thresholded by σ deviation of the trained sets) to the face part it is selected and added for new statistics, if not it is discarded; select only up to 20 features for each image.

5   Results and Evaluation

BioID [4] provides a free database of face images widely used to benchmark face identification and recognition algorithms. It has 1521 images with 23 different subjects in a variety of conditions such as talking, smiling, illumination changes, and small rotations. For the evaluation it was manually partitioned by us in 7 categories regarding frontal and rotation, open and closed eyes, smiling and neutral, wearing glasses. A set of 30 images was randomly picked from all the sets and were used as a training set. Results were averaged for 10 different rounds of images for training. The category sets are not uniform since the purpose of the manual partition was to analyze the particular variations on the database. Table.1 gives the exact numbers of images in the partitions, as well as the total features detected in the face parts and their average by image.

In Table.1 it can be seen that in all image partition sets the algorithm proposed detected 6 (30%, 6 out of 20 maximum) to 10 (50%, 10 out of 20 maximum), with average of 7 (35%, 7 out of 20 maximum) facial features, or face parts. As the images

from the database show relevant backgrounds, and the faces are under variations of lighting and expressions, by selecting 30% to 50% of relevant points in the face parts is a successful achievement for the task since there are only 4 face parts (eye left, eye right, nose, mouth) considered, and no training on templates or appearance models was done, but a fast feature selection and relative weighting on clustering was proposed. Similarly testes were also performed varying the size of the training set for randomly 10% (152) of the images, 30% (456) and 50% (760), and on average the number of detected features per image in the face parts (out of 20 maximum) was   respectively 7, 7, and 7, keeping the relevance around 35% as shown in Table.1

Table 1. Results (Cross-validated) showing the number of facial features in sets of   images from the BioID database. A random set of images (10%, 30% and 50%) was separated for training,. The maximum number of features that could be selected   is 20.

Partition set	Number of images used for testing			Total number of detected features in the face parts (ORB + SURF)			Average number of detected features per image in the face parts (out of 20 maximum)		
Number of images used for training	152 10%	456 30%	760 50%	152 10%	456 30%	760 50%	152 10%	456 30%	760 50%
expressions1	17	11	8	109	60	36	6	5	5
expressions2	59	47	34	541	454	327	9	10	10
expressions3	196	156	105	1199	949	627	6	6	6
expressions4	432	342	247	3509	2752	1976	8	8	8
normal1	355	279	200	2619	2065	1507	7	7	8
normal2	234	174	118	1303	1003	684	6	6	6
eyes1	76	56	49	615	456	347	8	8	7
Total	**1369**	**1065**	**761**	**9895**	**7739**	**5561**	**7**	**7**	**7**

Figure.3 shows some output images from the algorithm proposed. The images are taken from different partition sets and they show the selected (out of 20 maximum) feature points (mixed ORB and SURF) and the marked for reference face parts. Only points in the face parts would be 100% success. The variations on illumination, face rotation and expressions are illustrated. It can be seen that the selected features are concentrated on the faces mostly, and we know that these feature detectors would respond strongly to salient regions in the background. However, the concentration on the face, and especially on the face parts is the result of the proposed algorithm which positively weights points in the face parts and negatively outside. The results on the BioID database confirms that the proposed algorithm selects a subset of features from ORB and SURF localized mostly in the face, and in the face parts. Also, it has been shown that ORB detector responds much better to face features than the others SIFT and SURF, and it can be further explored for identity recognition as well.

The preference for the ORB and SURF features for this work were twofold: first, they were not explored fully yet for face parts, or face identification, as the SIFT detector [3] [12]; second their properties of robustness to noise, low computational complexity, and localization [15] [10] would favor their use instead of SIFT. Although on average the contribution of SURF features in the selected set of points in the face parts are around 5% for this database, their computation is one order of magnitude faster than SIFT [15], and SURF features would possibly bring more robustness to scale variations than ORB for different (although not tested here) databases. The presence of SURF features is constant and important on the final selected features. Matching would be done for ORB and SURF features as originally proposed [1] [15], although the results presented here open the path for exploring multiple face parts localization and hybrid matching schemes for binary and real valued features possibly.
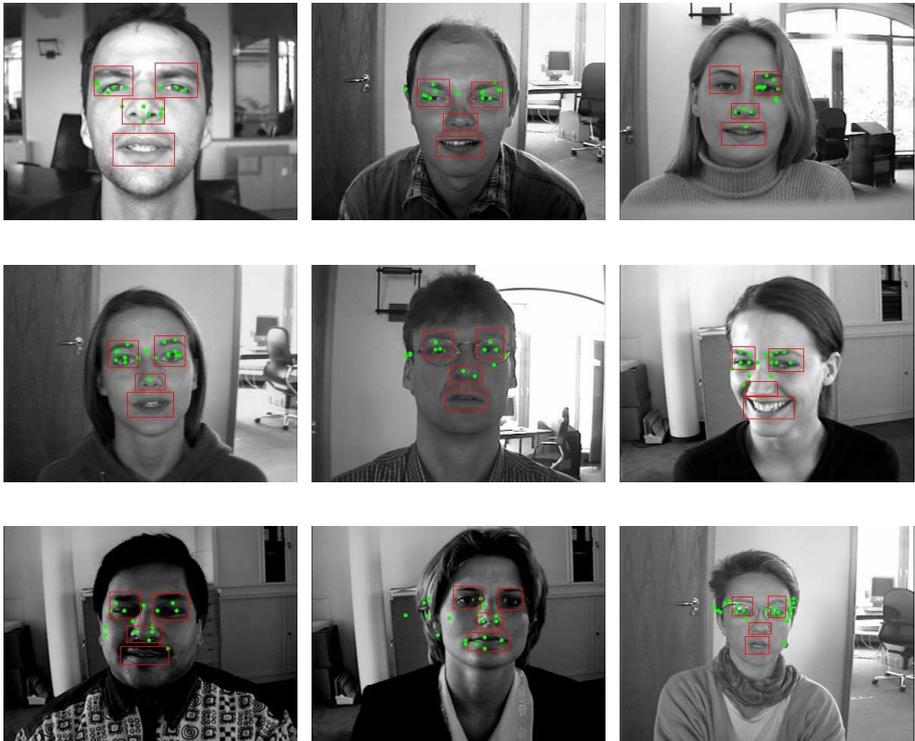
Fig. 3. Output images showing the selected features and their positions relative to the face parts

6   Conclusion and Future Work

In this paper we have proposed a new algorithm to train and select a subset of ORB and SURF features for face parts localization. It has been shown that the selected features from the algorithm concentrates on the face, and the facial features as eyes, eyebrows,

nose and mouth. The selected features represent 40% on average the regions detected on BioID, a benchmark free image database. At least 30% of the points detected were in the face parts considering the most challenging partition set. ORB features have been demonstrated here to be well suited for face identification, and face parts localization. Results are interesting to investigate further the use of combined ORB and SURF features for multiple face parts identification and recognition for close to real-time applications since those features are one order of magnitude faster than state of the art feature detector as SIFT.

References

1. Bay, H., et al.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU) 110(3), 346–359 (2008)
2. Belhumeur, P., et al.: Localizing Parts of Faces Using a Consensus of Exemplars. In: Proceedings of IEEE CVPR 2011, pp. 545–552 (2011)
3. Bicego, M., et al.: On the use of SIFT features for face authentication. In: Proceedings of CVPRW 2006 (2006)
4. BioID Face Database, http://www.bioid.com/ (visited in June 2013)
5. Chevallier, L., et al.: Facial Landmarks Localization Estimation by Cascaded Boosted Regression. In: Proceedings of VISAPP 2013 (2013)
6. Dantone, M., et al.: Real-time facial feature detection using conditional regression forests. In: Proceedings of IEEE CVPR 2012, pp. 2578–2585 (2012)
7. Ding, L., Martinez, A.M.: Features versus Context. IEEE Trans. PAMI 32(11), 2022–2038 (2010)
8. Dreuw, P., et al.: SURF-Face: Face Recognition under Viewpoint Consistency Constraints. In: Proceedings of BMVC 2009 (2009)
9. Gourier, N., Hall, D., Crowley, J.: Facial Features Detection Robust to Pose, Illumination and Identity. In: Proceedings of IEEE SMC 2004, pp. 617–622 (2004)
10. Heinly, J., Dunn, E., Frahm, J.-M.: Comparative Evaluation of Binary Features. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 759–773. Springer, Heidelberg (2012)
11. Klare, B., Jain, A.K.: On a taxonomy of facial features. In: Proceedings IEEE BTAS 2010, pp. 1–8 (2010)
12. Križaj, J., Štruc, V., Pavešić, N.: Adaptation of SIFT features for robust face recognition. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010, Part I. LNCS, vol. 6111, pp. 394–404. Springer, Heidelberg (2010)
13. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol 60(2), 91–110 (2004)
14. Miksik, O., Mikolajczyk, K.: Evaluation of Local Detectors and Descriptors for Fast Feature Matching. In: Proceedings of ICPR 2012, pp. 2681–2684 (2012)
15. Rublee, E., et al.: ORB: an efficient alternative to SIFT or SURF. In: Proceedings of IEEE ICCV 2011, pp. 2564–2571 (2011)