Markerless Augmented Reality based on Local Binary Pattern

Youssef Hbali, Mohammed Sadgal and Abdelaziz EL Fazziki
Faculty of Sciences SEMLALIA, Cadi Ayyad University, Prince Moulay Abdellah Avenue, Marrakech, Morocco

Keywords: Local Binary Pattern, Augmented Reality, Machine Learning, Face Detection, Eyes Detection.

Abstract: Augmented reality is becoming the future of e-commerce, throw their mobile devices, customers have access to all kind of information, going from weather, news papers, shops and so on. Today's mobiles devices are so powerful to the point that they can be used as a platform of virtual try-on systems. Over this paper we present a virtual eye glasses try-on system based on augmented reality and LBP for face and eyes detection. The well-known machine learning Ada Boost algorithm is used for real time eyes tracking, the resulting face and eyes positions are continuously utilized to overlay the glasses model over the face. The system helps evaluating glasses before trying them in the store and makes possible the design of its own style.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Object detection is a fundamental part of many virtual try-on systems. A flexible eye glasses try-on system which can be executed on mobile devices requires an efficient and robust face and eyes detection. Object detectors techniques can be divided into two main categories (Hjelmsa and Lowb, 2001) : feature-based approach where human knowledge is used to extract explicit object features such as nose, mouth, and ears for a face detection. The second approach is the image-based approach, in this approach, the object detection problem is treated as binary pattern recognition problem to distinguish between face and non-face images, eye and non-eye images, etc. This approach is a holistic approach that uses machine learning to capture unique and implicit object features. Based on the classification strategy used in the design process, image-based approach is categorized into two subcategories: appearance-based approach and boosting-based approach.

Appearance-based approach category is considered as any image-based approach face detector that does not employ the boosting classification methods in it classification stage. However, other classification schemes are used such as neural networks (Rowley et al., 1998) (Roth et al., 2000), Support Vector Machines (SVM) (Osuna et al.,), Bayesian classifiers (Cootesa et al., 2002) (Jin et al., 2004), and so forth. All techniques in the appearance-based approach lack the ability to perform in real-time, and it takes an order of seconds to process an image.

The other image-based approach subcategory is the boosting-based approach, this approach started after the successful work of Viola and Jones (Viola and Jones, 2001) where high detection rate and high speed of processing (15 frames/second) using the Ada Boost (Adaptive Boosting) algorithm (Freund and Schapire, 1996) and cascade of classifiers were used. Boosting-based approach is considered as any image-based approach that uses the boosting algorithm in the classification stage.

Augmented reality is a term for a live direct or indirect view of a physical, real world environment whose elements are augmented by computer-generated sensory input, such as sound or graphics (Lu et al., 1999) (Shen et al., 2010). Augmentation is conventionally in real-time, so is the need for a robust eyes detection system that is capable of processing image rapidly and detecting eyes accurately in an arbitrary face image with invariance to pose, scale and lighting.

2 LOCAL BINARY PATTERN

The local binary pattern (LBP) is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. The original LBP operator labels the pixels of an image by thresholding the 3-by-3 neighborhood of each pixel with the center pixel value and considering the result as a binary number. The decimal result is the sum of,

Hbali Y., Sadgal M. and EL Fazziki A..
Markerless Augmented Reality based on Local Binary Pattern.
DOI: 10.5220/0004531201370141
In *Proceedings of the 10th International Conference on Signal Processing and Multimedia Applications and 10th International Conference on Wireless Information Networks and Systems* (SIGMAP-2013), pages 137-141
ISBN: 978-989-8565-74-7
Copyright © 2013 SCITEPRESS (Science and Technology Publications, Lda.)

the thresholds multiplied by their weights values, as it can be seen in Fig 1.

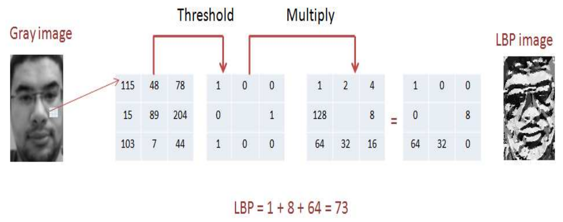

Figure 1: LBP Calculation.

In other words given a pixel position (x_c, y_c), LBP is defined as an ordered set of binary comparisons of pixels intensities between the central pixel and its surrounding pixels. The resulting label value of the 8-bit word can be expressed as follows :

$$LBP(x_c, y_c) = \sum_{n=0}^{7} t(l_n - l_c)2^n. \tag{1}$$

where l_c corresponds to the gray value of the central pixel, l_c the gray value of the neighbor pixel n, and function t(k) is defined as following :

$$t(k) = \begin{cases} 1, & \text{for } k \geq 0 \quad (2) \\ 0, & \text{for } k < 0 \quad (3) \end{cases}$$

According to (2), the LBP code is invariant to monotonic gray-scale transformations, hense the LBP representation may be less sensitive to illumination changes.
The 256-bin histogram of the labels computed over an image can be used as texture descriptor. Each bin of histogram (LBP code) can be regarded as micro-texton and the histogram characterizes occurrence statistics of simple texture primitive.
The histogram of the labeled image $f_l(x,y)$ can be defined as:

$$H_i = \sum_{x,y} I(f_l(x,y) = i), i = 0, .., L-1. \tag{4}$$

where L is the number of different labels produced by the LBP operator and l(A) is 1 if A true and 0 otherwise.

2.1 Multi-scale LBP

The LBP operator has been extended to consider different neighborhood sizes to deal with various scales (Ojala et al., 2002). The local neighborhood of the LBP operator is defined as set of sampling points equally spaced on a circle of radius R centered on the pixel to be labeled. These sampling points which do not fall exactly on the pixels are expressed using bilinear interpolation, therefore allowing any radius value and any number of points in the neighborhood. Fig. 3 shows different LBP neighborhoods.
The notation $LBP_{P,R}$ denotes the neighborhood of P sampling points on a circle of radius R.

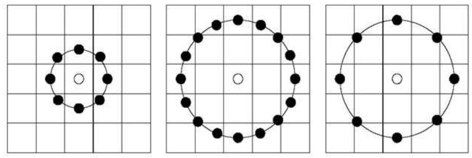

Figure 2: LBP operator examples : circular(8,1),(16,2) and (8,2).

2.2 Uniform Patterns

$LBP_{P,R}$ produces 2^P different binary patterns that can be formed by the P pixels in the neighbor set. Ojala et al. (Ojala et al., 2002) have noticed that most of the texture information was contained in a small subset of LBP patterns. Therefore, it is possible to use only a subset of the 2^P LBPs to describe the textured images. They defined these fundamental patterns as those with at most 2 bitwise transitions from 0 to 1 or vice versa. For example, 00000000 and 11111111 contain 0 transition while 0110000 and 01111110 contain 2 transitions and so on. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. For example, when using (8,R) neighborhood, there are a total of 256 patterns, 58 of which are uniform, which yields in 59 different labels.

3 LBP FACE FACIAL REPRESENTATION

Face image can be decomposed Hadid et al. (Hadid et al., 2004) introduced a face representation based on LBP for face recognition. To consider the shape information of faces, face images are divided into M small non-overlapping regions $R_0,R_1,...,R_M$ (as shown in Fig. 4). The LBP histograms extracted from each sub-region are then concatenated into a single, spatially enhanced feature histogram defined as :

$$H_{i,j} = \sum_{x,y} l(f_l(x,y) = i) l((x,y) \in R_j) \tag{5}$$

where i=,..L-1, j=0,..M-1. The extracted feature histogram describes the local texture and global shape of face images.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

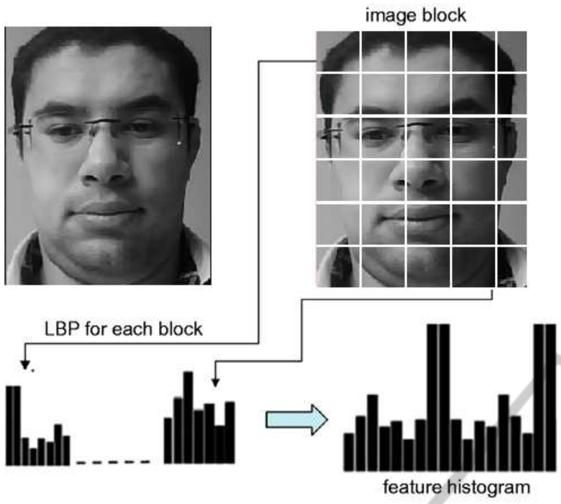

Figure 3: A face image is divided into sub-regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram.

4 AUGMENTED REALITY

Augmented Reality (AR) employs computer vision, image processing and computer graphics techniques to merge digital content into the real world. It enables real- time interaction between the user, real objects and virtual objects. AR can, for example, be used to embed 3D graphics into a video in such a way as if the virtual elements were part of the real environment.

Model-based tracking approaches(Reitmayr and Drummond, 2006) appear to be the most promising among the standard vision techniques currently applied in AR applications. While marker-based approaches such as ARToolkit(Kato and Billinghurst, 1999) or commercial tracking systems such as ART provide a robust and stable solution for controlled environments, it is not feasible to equip a larger outdoor space with fiducial markers. Hence, any such system has to rely on models of natural features such as architectural lines or feature points extracted from reference images.

For facial accessory products like eye glasses, it appears embarrassing to place such markers in front of the user who is trying on the glasses. Markerless AR systems use natural features instead of fiducial markers in order to perform tracking. Therefore, there are no ambient intrusive markers that are not really part of the world. Furthermore, markerless AR counts on specialized and robust trackers.

The first step of building the learning-based tracking system is to produce training data. For the proposed system we used two face databases, Bioid and CIE. From these different images we extract facial models under different rotations and both the left an right eyes. Then by applying a boosting (Friedman et al., 2000) classification we produce two classifier, one for faces, the other for eyes. In this system, a variant of Ada Boost, Gentle Ada Boost is used to select the feature and to train the classifier. The formal guarantees provided by the Ada Boost learning procedure are quite strong. It has been proved in (Freund and Schapire, 1996) approaches zero exponentially in the number of rounds. Gentle AdaBoost takes a newton steps for optimization.

The weak classifier is designed to select the single LBP histogram bin which best separates the positive and negative examples. Similar to (Viola and Jones, 2001), a weak classifier $h_j(x)$ consists of a feature f_j which corresponds to each LBP bin, a threshold θ_j and a parity p_j indicating the direction of the inequality sign:

$$h_j(k) = \begin{cases} 1 & \text{if } p_j f_j(x) \leq p_j \theta_j \quad (6) \\ 0 & \text{otherwise} \quad (7) \end{cases}$$

The boosted classifier is a combination of weights and weak classifiers.

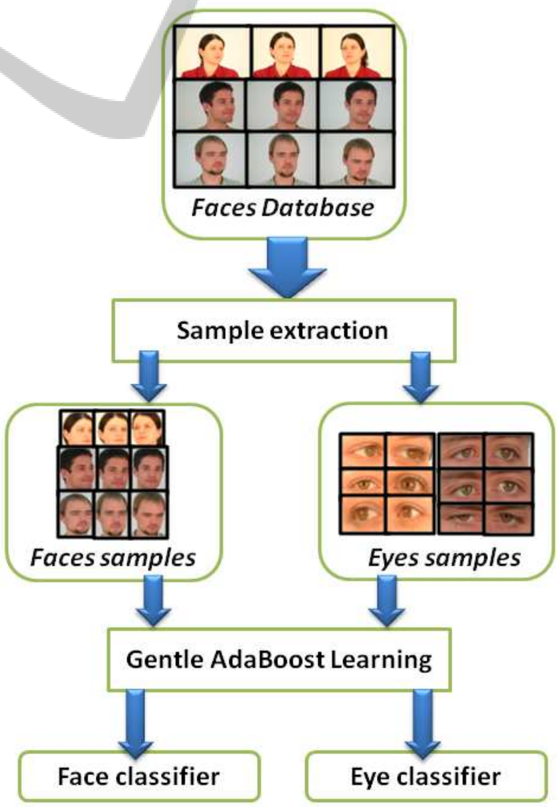

Figure 4: Off-line learning stage.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

5 EXPERIMENTAL RESULTS

In this section, we will measure the performance of the different classifier over Bioid database. The dataset of the Bioid database consists of 1521 gray level images with a resolution of 384x286 pixel. Each one shows the frontal view of a face of one out of 23 different test persons. For comparison reasons the set also contains manually set eye positions. The classifier takes a collection of marked up test images, applies the classifier and output the performance, i.e. number of found objects, number of missed objects and the number false alarms which are defined as following :

- Hits : the number of correctly found objects.
- Missed : the number of missed objects (must exist but are not found, also known as false negatives).
- False alarms : the number false alarms (must not exist but are found, also known as false positives).

The performance of the classifiers will be illustrated by the receiver operating characteristic (Laskoa et al., 2005) cuves which are frequently used in biomedical informatics research to evaluate classification and prediction models for decision support, diagnosis, and prognosis. ROC analysis investigates the accuracy of a model's ability to separate positive from negative cases (such as predicting the presence or absence of disease), and the results are independent of the prevalence of positive cases in the study population. It is especially useful in evaluating predictive models or other tests that produce output values over a continuous range, since it captures the trade-off between sensitivity and specificity over that range.

5.1 Test Performance for the Face Classifiers

For tests performance, we will compare the trained LBP classifiers to Haar feature classifiers, We apply both face classifiers on the Bioid database, the results are shown in the table 1.

Figure 5 gives the ROC curves comparing the performance of the two faces classifiers. It shows that there is little difference between the two classifiers in terms of accuracy.

Table 1 shows that LBP is much fast than the Haar classifier. The number of hits faces of LBP detector is higher than Haar detector and false alarms detected by Haar detector is considerable compared to the number of false alarms detected by the LBP detector.

Table 1: LBP and Haar faces classifiers performance evaluation and comparison.

Features	Hits	Missed	False alarm	Total time
LBP	1392	129	28	6.34 seconds
Haar	1377	144	547	157 seconds

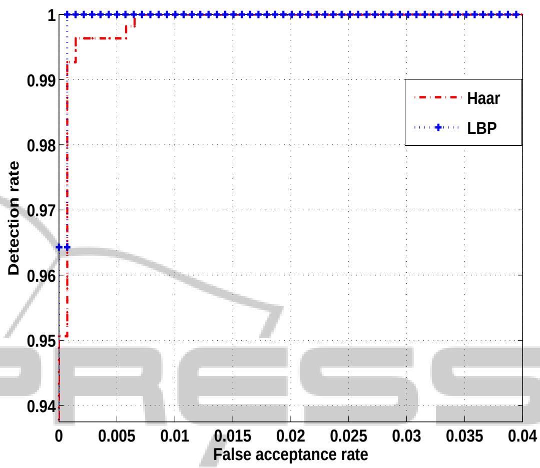

Figure 5: Roc curve of the LBP and Haar face classifier applied on the Bioid database.

5.2 Test Performance for the Eyes Classifiers

Table 2 shows the results of applying the eyes classifiers on the Bioid dataset.
It clearly appears the LBP eyes classifier gets down in terms of accuracy, however its shows that the LBP detector response time still more interesting than Haar detector response time.

Table 2: LBP and Haar eyes classifiers performance evaluation and comparison.

Features	Hits	Missed	False alarm	Total time
LBP	966	555	1993	43.6 seconds
Haar	1986	1056	630	73 seconds

6 PROPOSED ARCHITECTURE

For the augmented reality part, we apply the face classifier to first detect the face, see the figure 6.a . From this detected face we use the region of interest marked by the rectangle withing the face, see the figure 6.b, to look for eyes. Then, applying the eyes classifier to the region of interest leads to detect eyes withing the given image with better accuracy and better time response.

Finally, our system uses ARtoolkit to overlay a VRML (**V**irtual **R**eality **M**odeling **L**anguage) model

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

glasses on the image. Giving the user the ability to test many glasses without putting any kind of markers on his face.

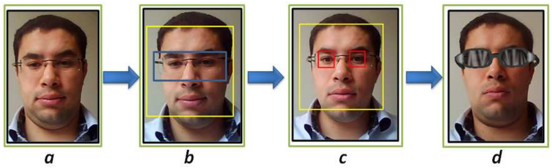

Figure 6: a) - Frame image rendered by the webcam, b) - The detected face and eyes region of interest, c) The detected eyes, d) The glasses model overlayed on the detected eyes.

7 CONCLUSIONS

In this paper, we introduced and implemented markerless augmented reality based on local binary patterns for eyes and face detection. The LBP features have proved accuracy on face detection, for small region like eyes, the LBP still need more improvements.
Due to the computational simplicity and speed of the LBP, virtual try-on systems can easily be implemented on mobile devices. This approach can improve highly the electronic commerce and will change the customers shopping habits.

REFERENCES

Cootesa, T., Wheelera, G., Walkerb, K., and Taylora, C. (2002). View-based active appearance models. *Image and Vision Computing*, 20(9-10):657664.

Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Proceedings of International Conference on Machine Learning*, page 148156.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of statistics*, 28(2):337–407.

Hadid, A., Pietikainen, M., and Ahonen, T. (2004). A discriminative feature space for detecting and recognizing faces. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2:797–804.

Hjelmsa, E. and Lowb, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, 3(83):236274.

Jin, H., Liu, Q., Lu, H., and Tong, X. (2004). Face detection using improved lbp under bayesian framework. *Multi-Agent Security and Survivability, 2004 IEEE First Symposium on*, pages 306–309.

Kato, H. and Billinghurst, M. (1999). Marker tracking and hmd cal-ibration for a video-based augmented reality conferencing system. *Augmented Reality, 1999. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94.

Laskoa, T., Bhagwatc, J., Kelly, H., and Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38(5):404–415.

Lu, S., Shpitalni, M., and Gadh, R. (1999). Virtual and augmented reality technologies for product realization. *CIRP Annals - Manufacturing Technology*, 2(2):471–495.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans, Pattern Analysis and Machine Intelligence*, 24(7):971–987.

Osuna, E., Freund, R., and Girosit, F. Training support vector machines: an application to face detection. *Computer Vision and Pattern Recognition, 1997. Proceedings IEEE Computer Society Conference on*, pages 130–136.

Reitmayr, G. and Drummond, T. (2006). Going out: robust model-based tracking for outdoor augmented reality. *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on*, pages 109–118.

Roth, D., Yang, M., and Ahuja, N. (2000). A snow-based face detector. *Advances in Neural Information Processing Systems*, pages 855–861.

Rowley, H. A., Baluja, S., and Kanade, T. (1998). Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1):23–38.

Shen, Y., Ong, S., and Nee, A. (2010). Augmented reality for collaborative product design and development. *Design Studies*, 31(2):118–145.

Viola, P. and Jones, M. (2001). Rapid object detecting using boosted cascade of simple features. *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1:511–518.

SCIENCE AND TECHNOLOGY PUBLICATIONS