Image Classifiers

Mikołaj Leszczuk, Remigiusz Baran, Michał Grega, Krzysztof Rusek, Piotr Guzik

Introduction

The IMCOP project—an "Intelligent Multimedia System for Web and IPTV Archiving. Digital Analysis and Documentation of Multimedia Content"—is a joint Polish-Israeli R&D project realized by a consortium of four partners. In general, IMCOP's objectives are twofold: multimedia data analysis and content discovery on one side, and data aggregation, content related binding (finding and assigning content related connections between data) and delivery on the other. Data analysis is performed in order to enrich the data (mainly images) by extracting their features and classifying them according to given criteria, as in Baran et al. (2015). A list of those criteria, as well as the classifiers developed so far in the IMCOP project, is as follows:

"Data analysis is performed in order to enrich the data (mainly images) by extracting their features and classify them according to given criterions"

- Age classifier
- Bokeh effect detector
- Dominant color of clothes (dresses)
- Dominant color counter
- Nudity identification
- People counter
- Profile/enface classification
- Red eyes detection
- Smile detection
- Unshaved faces

Further sections will explain the classifiers in detail.

Age Classifier

The age classifier decides whether the detected face belongs to a person that is older or younger than 18. The features for classification are histograms of local binary patterns (LBP). Support Vector Machine (SVM) with radial basis kernel (RBF) is used for classification. The algorithm returns both classification result and the probability of the indicated class.

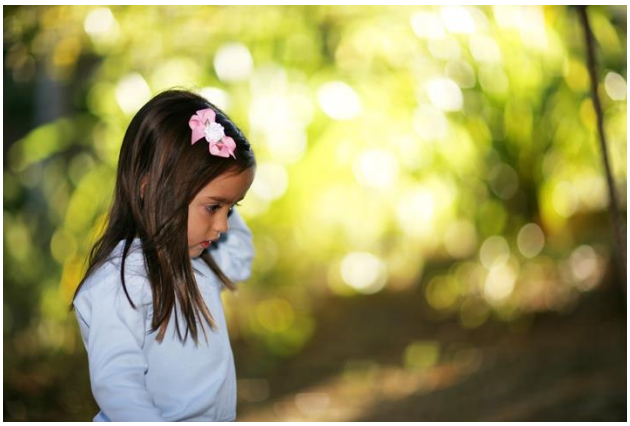

Figure 1. Coarse bokeh on a photo shot with an 85 mm lens and 70 mm **entrance pupil** diameter, which corresponds to *f*/1.2. "Josefina with Bokeh" by carlosluis – http://www.flickr.com/photos/paseodelsur/51805888/. Licensed under CC BY 2.0 via Commons – https://commons.wikimedia.org/wiki/File:Josefina_with_Bokeh.jpg#/media/File:Josefina_with_Bokeh.jpg.

Bokeh Effect Detector

In **photography**, **Bokeh** is the aesthetic quality of the blur produced in the out-of-focus parts of an image produced by a lens. Bokeh occurs for parts of the scene that lie outside the **depth of field** (Figure 1).

The algorithm, still a work in progress, is based on a combination of detection of out-of-focus blur (Leszczuk et al. (2014)) and detection of faces (Rusek et al. (2013), Rusek et al. (2014)).

The input is an RGB image of a face. The output is binary (Bokeh effect, no Bokeh effect) with confidence probability.

Dominant Color of Clothes (Dresses)

This classifier allows identification of the color of the dress of an actress in a so called "red carpet" photo. This scenario is useful for automated generation of content, when the designer

of the system wants to achieve high diversity of photos presented to the user.

Technically this solution is based on face detection. The largest face in the image is identified, which allows identification of the sampling region located in the hip area.

For purpose of clustering the color is quantized into one of nine categories: red, green, blue, black, white, magenta, yellow, cyan and colorful.

Dominant Color Counter

This classifier is capable of counting and identifying the dominant colors in the image. It is compliant with the Dominant Color descriptor as described in the MPEG-7 standard ISO/IEC 15938.

Nudity Identification

The algorithm checks whether the input image contains nudity or not. It is based on a statistical model of human skin color and some additional shape information obtained with use of discrete cosine transform (DCT). Support vector machine (SVM) with radial basis function (RBF) kernel is used for classification. The result is the probability that the image contains nudity.

People Counter

The people counter module is a simple wrapper around the face detector from OpenCV Library. Haar cascades are used for face detection, and the number of detected face regions is assumed to be the number of people in the image.

Profile/En Face Classification

Preliminary work on profile classification was published in (Rusek 2011). The current implementation is far more accurate compared to the previous statistical model of the face image.

In this new approach, the face is classified as facing left, right, or front. Support vector machines are used for classification. The features used for classification are the concatenated histograms of local binary patterns (LBP) calculated over a 10 × 10 pixel sliding window. Before feature extraction, the size of a face image is normalized to 100 × 100 pixels.

Red Eye Detection

Taking a picture in bad light condition often ends with the red eye effect. Despite many correcting algorithms, the effect is still possible. Red eye is detected by combining two algorithms. The first algorithm locates the positions of the eyes in the face image. The second algorithm, for each pixel, calculates the probability that the pixel belongs to red eye. Probability is calculated using statistical model of the red eye color based on generalized linear models. A decision is made, based on the average redness, which is the average probability of a red pixel, weighted by the distance to located eye pupils.

Smile Detection

Research on smile detection has gained a lot of attention. Our approach relies on face detection in grayscale images. For each detected face region, smile is detected by support vector machines. The features used in the algorithm are the mean and standard deviation of coefficients of Haar wavelets decomposition of the image. These are the well-known texture descriptors. Such descriptors are collected over three scales ranging from an 8 × 8 to a 64 × 64 pixel sliding window. Before

feature extraction, face region in the base image size is rescaled to 128 × 128 pixels.

Unshaved Faces

Detection of unshaved faces can be a complicated task because of the many styles of beard and moustache. In IMCOP, unshaved faces are recognized by support vector machines. The image is pre-processed by calculating skin probability for each pixel. The skin colour model is a Generalized Linear Model of RGB components. A nonlinear model up to third power is constructed. SVM input is the raw value of probability map scaled to 10 × 10 pixels.

Contribution to VIME

As described in the previous article, the VIME (Video and Image Models for consumer content Evaluation) project in VQEG (Video Quality Experts Group) is dedicated to No-Reference image quality and is currently collecting images to build a dataset that can be used to conduct our research work. VIME has set up a Flickr group where contributors can upload images to this dataset.

The indicators can contribute to VIME by automatically adding Flickr **machine tags**. A machine tag or **triple tag** uses a special **syntax** to define extra **semantic** information about the tag, making it easier or more meaningful for interpretation by a computer program. Machine tags comprise three parts: a **namespace**, a **predicate**, and a **value**.

References

[1] Baran, R., Zeja, A., Ślusarczyk, P., "An overview of the IMCOP system architecture with selected intelligent utilities emphasized," *Proceedings of Eight International Conference on*

Multimedia Communications, Services and Security (MCSS 2015), Nov. 2015.

[2] Leszczuk, M., Hanusiak, M., Farias, M. C. Q., Wyckens, E., Heston, G. "Recent developments in visual quality monitoring by key performance indicators.," *Springer Multimedia Tools and Applications*, pp. 1-23, Sept. 2014. DOI: 10.1007/s11042-014-2229-2.

[3] Rusek, K., Orzechowski, T., Dziech, A., "LDA for face profile detection," *Proceedings of Fourth International Conference on Multimedia Communications, Services and Security (MCSS 2011)*, 2011.

[4] Rusek, K., Guzik, P. "Neural network regression of eyes location in face images," *Proceedings of Sixth International Conference on Multimedia Communications, Services and Security (MCSS 2013)*, June 2013.

[5] Rusek, K., Guzik, P. "Two-Stage Neural Network Regression of Eye Location in Face Images," *Springer Multimedia Tools and Applications*, pp. 1-14, July 2014. DOI: 10.1007/s11042-014-2114-z.

Mikołaj Leszczuk
AGH University of Science and Technology, Poland

Remigiusz Baran
Kielce University of Technology, Poland

Michał Grega
AGH University of Science and Technology, Poland

Krzysztof Rusek
AGH University of Science and Technology, Poland

Piotr Guzik
AGH University of Science and Technology, Poland