



Comparison Of Viola-Jones and Kanade-Lucas-Tomasi Face Detection Algorithms

KAMATH AASHISH and A VIJAYALAKSHMI

Department of Computer Science, Christ University, Bengaluru, India.

*Corresponding author E-mail- kamath.aashish@mca.christuniversity.in

<http://dx.doi.org/10.13005/ojcs/10.01.20>

(Received: March 04, 2017; Accepted: March 09, 2017)

ABSTRACT

Face detection technologies are used in a large variety of applications like advertising, entertainment, video coding, digital cameras, CCTV surveillance and even in military use. It is especially crucial in face recognition systems. You can't recognise faces that you can't detect, right? But a single face detection algorithm won't work in the same way in every situation. It all comes down to how the algorithm works. For example, the Kanade-Lucas-Tomasi algorithm makes use of spatial common intensity transformation to direct the deep search for the position that shows the best match. It is much faster than other traditional techniques for checking far fewer potential matches between pictures. Similarly, another common face detection algorithm is the Viola-Jones algorithm that is the most widely used face detection algorithm. It is used in most digital cameras and mobile phones to detect faces. It uses cascades to detect edges like the nose, the ears etc. However, if there is a group of people and their faces are close to each other, the algorithm might not work that well as edges tend to overlap in a crowd. It might not detect individual faces. Therefore, in this work, we test both the Viola-Jones and the Kanade-Lucas-Tomasi algorithm for each image to find out which algorithm works best in which scenario.

Keywords: Viola-Jones, Kanade-Lucas-Tomasi, face detection, face recognition.

INTRODUCTION

Face detection is a computer technology that identifies human faces in digital pictures. Face detection may also refer to the mental process by which humans locate visual scenes in real life. Face detection is a particular case of class-object

detection. In class-object detection, the main task is to look for the sizes and locations of all objects in a picture that are from a given class. Some examples include torsos, cars, and pedestrians. Face-detection algorithms focus on finding human faces.^[1]

An effective approach for detecting faces is based on the algorithms which are genetic and the eigen-face technique:

At first, all the valley type regions in the level grey image are checked for regions of possible human eyes. Then this genetic algorithm is used to create all possible regions of the face which include the iris, the mouth corners, the eyebrows and the iris. All possible candidates of the face are normalised to remove effects of lighting caused due to inconsistent illumination and the effect of shirring due to movement of the head. The value fitness of each candidate is checked based on its showing on the eigen-faces. After a lot of iterations, all candidates of the faces with a great value fitness are given for more verification. At this stage, the symmetry of the face is checked and the existence of so many facial features is checked for all face candidates.^[1]

Detection of faces is gaining marketers' interest. A webcam can be plugged into a television set and find any face coming around. The system then measures the age, gender, and race range of the faces. Once the data is gathered, a chain of ads can be shown that is specific to the detected age/gender/race. A great example of one kind of a system is OptimEyes and is embedded into the Amscreen digital sign system. A recent innovation in the field was by AdMobilize based in Miami Beach. AdBeacon, became the world's first 'device to plug in and check real time analytics', allowing any retailer to get the exact same face detecting tech that large advertisers use. They've also created the phrase pay-per-face. Some other examples are SDK of Visage Technologies which gives face detection information to researchers who are into marketing, including analytics such as emotion detection, gender recognition and age finding and freeware software such as OpenFace.^[1]

The face detection of humans is not really about recognition of faces – it is more about detection of faces. It has been known that the step 1 in detecting faces automatically – the accurate finding of faces of humans in random scenes, is the most important process in there. When faces can be found exactly in several scenes, the step of recognition after that is not so complex now. So

this paper tries to collect all data available about process of face detection automatically.^[2]

Recently, recognition of faces has gained attention and its research has quickly grown by engineers and neuroscientists, since it has many practical uses in communication of computer vision and accessing control systems automatically. Especially, detection of faces is a critical part of recognition of faces as the step 1 of face recognition automatically. However, face detection is not a plain idea because it has a lot of changes of picture appearance, such as difference in poses (front, non-front), image orientation, occlusion, facial expressions and illuminating conditions.^[3]

Many unique methods have been told to solve each difference listed above. For example, the methods to match templates are used for localising faces and detection by calculating the relation of an input picture to a standard pattern of faces. The feature variation approaches are used for detecting features of mouth, ears, eyes, nose, etc. The appearance-based methods are necessary for detecting faces with neural network of eigen faces and approach of data that is theoretical. Anyway, practicing the methods listed is still a great obstacle. Luckily, the pictures used in here have a uniformity degree and so the algorithm for detection can be simpler: all the faces have front view and are vertical; also, they are under the same illumination condition. Skin colour finding in colourful pictures is a famous technique for detecting faces. Many ways have shown how to locate regions of skin colours in the input picture. While the input colourful picture is usually in the RGB format, these methods always use colour components in the colourful space, such as the YIQ or HSV formats. That is because components of RGB are subject to the conditions of lighting thus the detection of faces may fail if the conditions of lighting changes.^[3]

The faces of humans plays a critical role in our interaction socially, showing identity of people. Using the faces of humans as a security key, face recognition with biometric technology has gained great amount of attention in the past several years due to its ability for a large variety of use in both law enforcement and non-law enforcement. As

compared with some other available biometrics systems using palm prints/finger prints and iris checking, face recognition has some advantages due to its non-contact process. Facial pictures can be taken from a large distance without being in contact with the person being checked, and the identification does not need social interaction with the person. Moreover, face recognition serves the purpose of deterring crime because facial pictures that have been recorded and saved can later help identify a criminal.^[4]

In this paper, we will test the Viola Jones algorithm in different situations to detect its accuracy which will vary according to the situations' brightness, number of faces, lighting, movement speed etc. We will analyse how well it works in each situation and find situations where it might be wise to consider using some other algorithm for better face detection. We repeat the same procedure for the Kanade-Lucas-Tomasi algorithm to know the strengths and weaknesses of each algorithm. To compare the algorithms, we used several images from different sources. We classified them into five categories - front face, looking left, looking right, looking up and looking down. The first three categories were further divided into four subcategories - bright, very bright, dark and very dark. We ran each image through both the Viola-Jones and the Kanade-Lucas-Tomasi algorithm to check their detection rates.

Viola-Jones detected 87% of the total faces whereas Kanade-Lucas-Tomasi detected only 84%. An interesting thing to note is that out of all the images that were run through both the algorithms, Viola-Jones detected faces in a few images that weren't detected by the Kanade-Lucas-Tomasi algorithm but there were no images whose faces

were detected by Kanade-Lucas-Tomasi but not by Viola-Jones.

Concept

Concepts overview

Face detection is the method of locating faces of humans automatically in digital media such as pictures and videos. A detected face is reported at a particular position with an related orientation and size. A human face can be searched for critical landmarks such as the nose and eyes after being detected.^[5]

Here are some of the jargon that we use in face detection discussion:

Face recognition on its own decides if two faces probably belong to the same human. But Google Face API provides functionality limited to detection of faces and not recognition of faces. Tracking of faces extends detection of faces to sequences video across video frames. Any face of a person appearing in a video for any amount of time can be tracked automatically. That is, human faces that are found in continuous video frames can be decided as being the same human. Know that this is not a form of recognition of faces; this method just makes conclusions based on the motion and position of the faces in a continuous video sequence.^[5]

A point of interest shown on a face is a facial landmark. Examples are the right eye, left eye, and base of nose. The great ability to find landmarks on a face that is found is given by the Face API.^[5]

This API finds the entire face regardless of landmark information in detail instead of initially finding landmarks and using them as a basis for whole face detection. Thus detection of landmarks is



Fig. 1: Facial landmarks^[5]

an optional way that can be done after the detection of a face. Since it takes more time to process, detection of landmarks is not done on its own. You can optionally mention that detection of landmarks should be completed. Determining whether a particular characteristic of the face is present is called classification. Eg. We can classify a face with whether it has open or closed eyes. Another particular example is whether it's a smiling face or not.^[5]

Face orientation

Fig. 2. Pose angle estimation. (a) The coordinate system with the image in the XY plane and the Z axis coming out of the figure. (b) Pose angle examples where y =Euler Y, r =Euler Z.^[5]

The Euler X, Euler Y, and Euler Z angles characterise a face's orientation as shown in Fig. 2. The Face API provides measurement of Euler Y and Euler Z (but not Euler X) for detected faces.^[5]

The facial angle Euler Z is always shown. The facial angle Euler Y is found only while using the special setting of "accurate" mode face detector (because the "fast" mode setting, skips some processes for faster detection).^[5]

Classification

Determining whether a particular facial characteristic exists is called classification. For now, two classifications are detected by the Android API: smiling and open eyes. The iOS Face API right now detects the smiling classification. Classification is

displayed as a value of certainty, signalling the confidence that the characteristic of the face exists. Eg. a value of 0.7 or greater for the classification of smiling shows that it is likely that it's a smiling person.^[5]

Both of these classifications depend heavily on detection of landmarks.^[5] Also note that "open eyes" and "smiling" classifications only work for front side faces, that is, faces with a tiny Euler Y angle (max +/- 18 degrees).^[5]

LITERATURE SURVEY

In⁹, we see that if an image of a random size contains a face of a person, it must be known by a face detector. One way to solve the problem is using classification of binary which has a particular classifier is made to reduce the risk of misclassification. Since we can't know the real previous probability for a particular picture to have a face, in order to achieve an acceptable performance, the specific algorithm must reduce both the false negative and positive rates.

This goal needs a specific arithmetic set of what differentiates faces of people apart from other things. These features can be known with Adaboost, a new committee learning algorithm which depends on a group of classifiers that are too weak to form a stronger thing through a mechanism of voting. Generally, if a classifier is too weak, it can't meet a previously set target of classification in terms of errors.

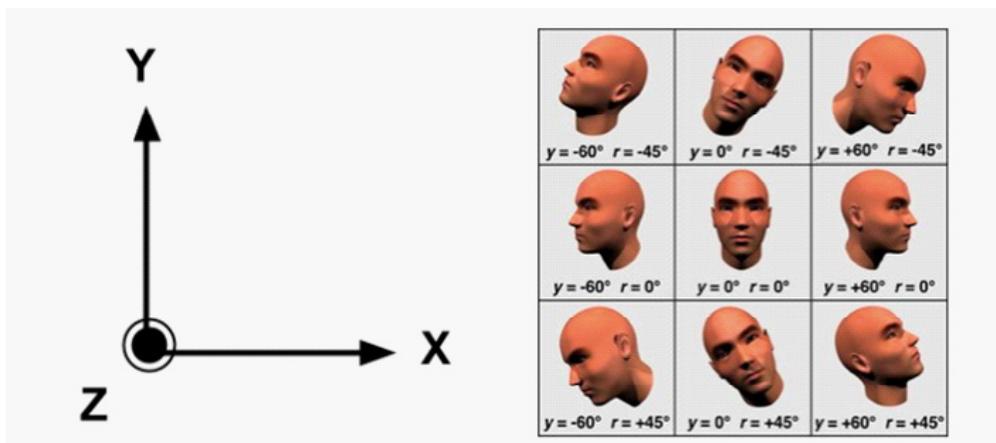


Figure 2. Facial orientation ^[5]

It's important that an algorithm of operations have a sufficient budget for this task. Some of the methods such as attentional cascade and integral image make the Viola-Jones algorithm very efficient: it performs sufficiently good on a standard PC if it's given real time image sequences generated from ordinary webcams.

In¹⁰, the author describes a machine learning approach for detecting visual objects which has the capacity to process different images very quickly and achieving great detection rates. This is possible due to three key contributions. The main one is a representation of an image called "Integral Image" which lets very fast computation of the detector's features. The second is an Adaboost - based continuous algorithm of learning which chooses a tiny number of important visual features from a larger set and gives classifiers that are very efficient.

In¹¹, we know that detecting human faces has been a critical issue for gesture, expression and face recognition for a long time. Although several attempts were made to localise and detect faces, these approaches assumed things that led to restriction of their applications to generic cases. They identified that the critical factor in robust and generic systems is that of using a huge quantity of picture proof related by knowledge of detailed models through a framework of probability.

The authors show an algorithm based on features for face detection that is quite general and is also very extensible to catch up with much more heavy differences in the imaging conditions. The algorithm uses points of features detection from images using laters and classifies them into facial candidates using grey and geometric constraint levels. To check the possibility of any facial candidate and to reinforce probabilities, a probabilistic framework is used.

They provide sufficient results to support the approach's validity and show its capability of face detection under different viewpoints, orientations and scales.

In¹², they show a new locating human faces method in a background that is complicated. There are three levels in this hierarchical knowledge based

method system. At varying resolutions and focused on mosaic images, two pictures are selected as higher. An upgraded method of detecting edges is shown in the lower level. Unknown human faces can be located spanning a vast array of sizes in a black-white picture that is complicated.

In¹³, the author presents a vast range of object classification/recognition methods based on moments of images. They review several types of moments such as complex moments, geometric moments and invariants based on moments with respect to different image distortions (scaling, rotation, image blurring, affine transform etc.) and degradations which may be able to be used as classification shape descriptors.

This explains a common idea on how to construct some of these invariants and in the process show a few of them in explicit forms. They check up effective algorithms that can be used for computing moments and show practical examples of using moment invariants in practical world applications.

In¹⁴, they're developing a method of object detection combining bottom-up image segmentation with top-down recognition. The two important ways in this method: a step to check and a step to generate a particular hypothesis. An upgraded feature to context shape which is much more efficient with background clutter and object deformation is designed in the top-bottom step of generating hypothesis.

The Shape Context which has been improved is used to generate a new hypotheses set of figure-ground masks and object locations which have low precision and high recall rate. In the step of verification, they check a segmentations set that are feasible and in line with top-bottom hypothesis object. Then we use a method of FPP (False Positive Pruning) to remove positives that are false.

They take advantage that positive regions that are false usually do not fall in line with any ordinary segmentation of images that are feasible. Several experiments show that this ordinary framework has the capacity to achieve both high precision and high recall with just a small amount

of training examples that are positive and this way can be generalised to several different classes of objects.

In¹⁵, they try to review a vast range of techniques used for face detection in detail. This includes ICA, LDA, SVM, PCA and Gabor wavelet tool for soft computing like ANN for detection and different hybrid combinations of these methods. This review checks all the techniques with parameters that question face detection like pose variation, illumination and facial expressions.

In¹⁶, the midline symmetry which is crucial of faces of humans is proved to have an important part in facial detection and coding. This even has crucial and significant relations with finding of the primate cortex organisation as well as experiments on human psychophysiology. There's significant evidence that the facial dimension detection space for faces of humans is much lower than estimated previously.

An outcome of the present method is the making of a distribution of probability in facial space that creates a realistic and interesting range of faces. One more outcome is a 100% accuracy detection algorithm when measured by reasonable criteria.

In¹⁷, they show a different approach to checking shapes similarity and take advantage of it for object detection. In this framework of theirs, the similarity measurement is followed by:

1. Checking for connections between different points on two different shapes.
2. Using those connections to get an estimate an transformation of aligning.

They put a shape context and a descriptor to each point. A reference point's shape context finds out the distribution of the leftover relative points thus offering a discriminative characterisation which is global.

Data Collection and Preprocessing

In this paper, we will test the Viola Jones algorithm in different situations to detect its accuracy which will vary according to the situations' brightness, number of faces, lighting, movement

speed etc. We will analyse how well it works in each situation and find situations where it might be wise to consider using some other algorithm for better face detection. We repeat the same procedure for the Kanade-Lucas-Tomasi algorithm to know the strengths and weaknesses of each algorithm.

To compare the algorithms, we used several images from different sources. we classified them into five categories - front face, looking left, looking right, looking up and looking down. The first three categories were further divided into four subcategories - bright, very bright, dark and very dark.

Viola-Jones detected 87% of the total faces whereas Kanade-Lucas-Tomasi detected only 45%. An interesting thing to note is that out of all the images that were run through both the algorithms, Viola-Jones detected faces in five images that weren't detected by the Kanade-Lucas-Tomasi algorithm but there were no images whose faces were detected by Kanade-Lucas-Tomasi but not by Viola-Jones.

Methodology and implementation

Methodology

Viola - jones

There are four stages: Haar selection of characteristics, creation of images that are integral, training of adaboost, classifiers that cascade.

Human faces have some common characteristics. For example, eye regions are darker than cheek regions. Also, nose region is brighter than eyes.^[6]

A representation of pictures called the integral image checks rectangular features in real time, which gives them a great edge of speed over more complex alternate features. Because every feature's rectangular area is usually next to to at least one more rectangular feature, it implies that any two-rectangle characteristic can be calculated in six references of arrays, any three-rectangle characteristic in eight, and any four-rectangle characteristic in nine.^[6]

In cascading classifiers, faces consist of only 0.01% of sub windows. It's a waste of time

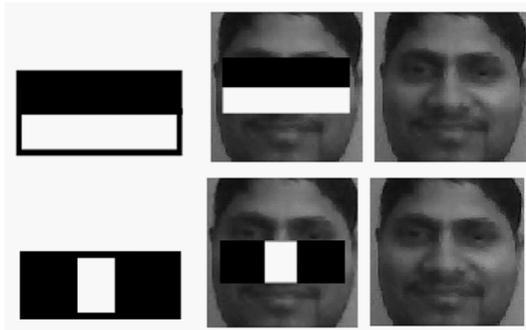


Fig. 3: Feature detection^[6]

to compute negative sub windows, so we should waste time only on sub windows that are positive. To achieve this, we can use a 2-feature classifier. The first one can act as a first line of defence to remove all negative sub windows. The second one can be used to remove the negatives that were harder to detect in the first layer etc. Gradually much more complicated classifiers cascade achieving better rates of detection.^[6]

Viola-Jones has several advantages like feature selection that is sophisticated and an invariant detector that locates scales. We can scale the features instead of scaling the image itself. Since it's a general scheme of detection, it can be trained for detecting other things like cars.^[6]

But Viola-Jones comes with disadvantages too. It is not as effective detecting tilted or turned faces. It is sensitive to lighting conditions and there could possibly be different detections of the exact face due to sub windows overlapping.^[6]

Kanade - lucas - tomasi

It is shown mostly for dealing with the question that traditional techniques of image

Table 1: Overall results of face detection by both algorithms

	Viola-Jones	Kanade-Lucas-Tomasi
Looking front	97%	90%
Looking left	90%	85%
Looking right	88%	83%
Looking up	80%	80%
Looking down	80%	80%
Total	87%	84%

registration are usually expensive. Kanade-Lucas-Tomasi uses data on spatial intensity to guide searching for the position yielding the most accurate match. It is much quicker than ordinary methods for checking lesser probable matches between the pictures.^[7]

Implementation

The test face images are tested through both of the algorithms. In some instances, Viola-Jones detected the face, but Kanade-Lucas-Tomasi didn't. In some cases, neither detected the faces. In majority of the cases, both detected the faces. But there was no instance where Kanade-Lucas-Tomasi detected the faces but Viola-Jones didn't.

CONCLUSION

Viola-Jones has a great detection rate in every scenario and is better than the Kanade-Lucas-Tomasi in every scenario.

Examples:

Both Viola-Jones & Kanade-Lucas-Tomasi detect:

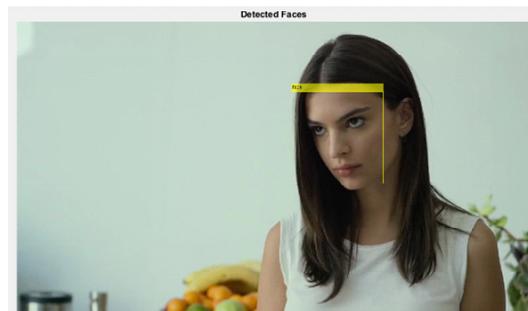
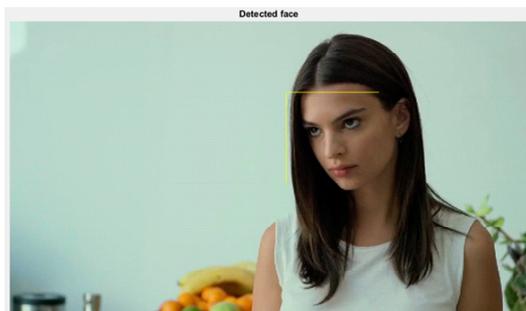


Fig. 5: Face detection by both Viola-Jones & Kanade-Lucas-Tomasi.^[18]

Neither Viola-Jones nor Kanade-Lucas-Tomasi detect:

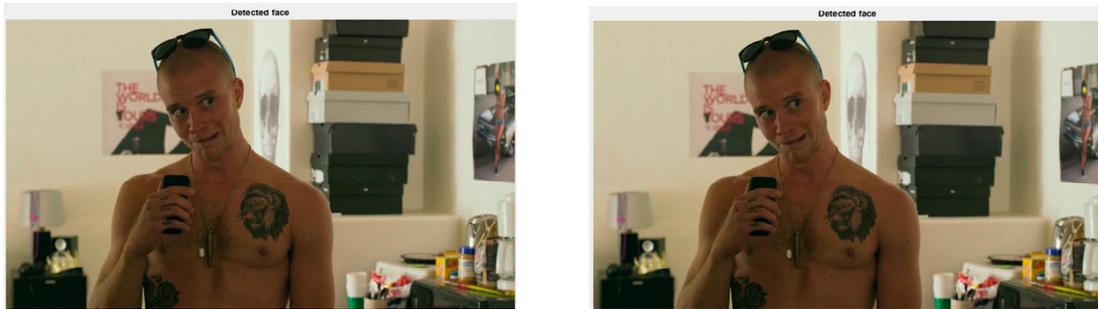


Fig. 6: Face detection by neither Viola-Jones nor Kanade-Lucas-Tomasi^[18]

Viola-Jones detects but Kanade-Lucas-Tomasi doesn't:

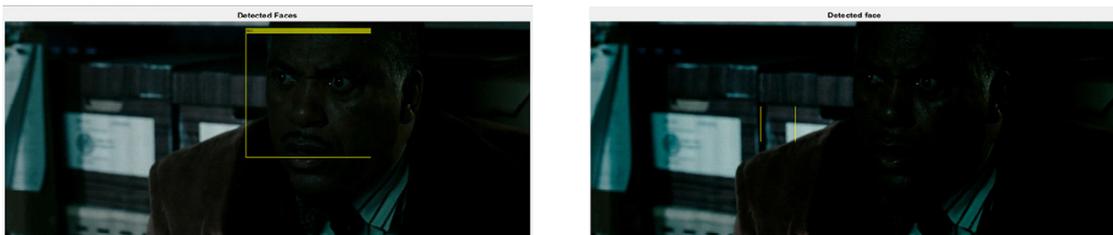


Fig. 7: Face detection by Viola-Jones but not by Kanade-Lucas-Tomasi^[19]

Kanade-Lucas-Tomasi detects but Viola-Jones doesn't:

(no such scenario)

REFERENCES

1. Face detection - Wikipedia - https://en.wikipedia.org/wiki/Face_detection
2. Face detection - facedetection.com.
3. Inseong Kim, Joon Hyung Shim and Jinkyu Yang (2016) *Face Detection*, Stanford University, *International Journal of Engineering Research and Applications*, 6, 1, pp145-150.
4. Face Recognition - nec.com.
5. Face detection concepts overview - Google Developers - <https://developers.google.com/vision/face-detection-concepts>.
6. Viola-Jones object detection framework - Wikipedia - https://en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_framework.
7. Kanade-Lucas-Tomasi feature tracker - Wikipedia - https://en.wikipedia.org/wiki/Kanade-Lucas-Tomasi_feature_tracker.
8. Paul Viola, Michael Jones, 'Robust Real Time Face Detection', *International Journal of Computer Vision*, 57(2), 137 - 154.
9. Yi Qing Wang, 'An Analysis Of The Viola - Jones Face Detection Algorithm', *Image Processing On Line*, 2014, v0.5.
10. Paul Viola, Michael Jones, 'Rapid Object Detection using a Boosted Cascade of Simple Features', *Accepted Conference on Computer Vision and Pattern Recognition*, 2001.
11. Kin Choong Yow, Roberto Cipolla, 'Feature based Human Face Detection', *CUED, F-INFENG, TR249*.
12. Guangzheng Yang, Thomas Huang, 'Human Face Detection In A Scene', *IEEE*, 1063-

- 6919/93.
13. Jan Flusser, 'Moment Invariants In Image Analysis', World Academy Of Science, *Engineering and Technology*, **11**, 2005.
 14. Liming Wang, Jianbo Shi, Gang Song, 'Object Detection Combining Recognition and Segmentation'.
 15. Sujata Bhele, Vijay Mankar, 'A Review Paper On Face Recognition Techniques', *International Journal Of Advanced Research in Computer Engineering and Technology*, **1**(8), 2012.
 16. Lawrence Sirovich, Marsha Meytlis, 'Symmetry, Probability And Recognition In Face Space', *PNAS*, **106**(17), 6895 - 6899.
 17. Serge Belongie, Jitendra Malik, Jan Puzicha, 'Shape Matching And Object Recognition Using Shape Contexts', *IEEE Transactions On Pattern Analysis And Machine Intelligence*, **24**(24), 2002, 509 - 522.
 18. Scene from the movie 'We Are Your Friends'.
 19. Scene from the movie 'Watchmen'.