ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2011. No. 10(116)
ELEKTRONIKA IR ELEKTROTECHNIKA

SYSTEM ENGINEERING, COMPUTER TECHNOLOGY
T 120
SISTEMŲ INŽINERIJA, KOMPIUTERINĖS TECHNOLOGIJOS

Inspection System based on Computer Vision

M. Petkevicius
Department of Multimedia Engineering, Kaunas University of Technology,
Studentų str. 50-401, LT-51367 Kaunas, Lithuania, phone: +370 37 300371, e-mail: martynas.petkevicius@stud.ktu.lt
A. Vegys
Software Engineering Department, Kaunas University of Technology,
Studentų str. 50-406, LT-51367 Kaunas, Lithuania, phone: +370 37 300361, e-mail: andrius.vegys@stud.ktu.lt
T. Proscevicius, A. Lipnickas
Department of Process Control Technology, Kaunas University of Technology,
Studentų str. 48-111, LT-51367 Kaunas, Lithuania, phone: +370 602 10302, e-mail: tomas.proscevicius@gmail.com

crossref http://dx.doi.org/10.5755/j01.eee.116.10.889

Introduction

Advances in computer science and robotics make automated surveillance systems possible. Mobile inspection robots or stationary inspection cameras make use of visual information to detect people and their suspicious activities. Photographs of potential offenders can then be taken and human security officers informed to take a closer look.

Instead of using only human silhouette detection for this task, it was decided to combine it with face detection. Not only this would make human detection more accurate in cases where only part of silhouette is visible, but it would also provide with additional information, such as which side the observed person is facing, and would make face recognition [1] possible.

Human silhouette detection

For human silhouette detection our system employs histograms of oriented gradients (HoG) [2] method. The first step in HoG algorithm is to calculate gradients of each image pixel. This can be achieved using simple filter kernels [-1, 0, 1] and $[-1, 0, 1]^T$. Image window is then divided into cells. For each cell a gradient orientation histogram is calculated. Cells are then grouped into larger units called blocks. Blocks overlap, so each cell contributes to more then one block. Histograms within blocks are then normalized to account for variance in illumination and contrast. Block histograms then form the final descriptor. Classifier is learnt using support vector machines (SVM) [3].

In our experiments, HoG implementation found in "OpenCV" computer vision library was used. Default classifier was employed.

Face detection

For face detection it was decided to use Haar training [4] because of its high detection speed. This method uses different kind of image representation called integral image. Value of specific pixel of integral image is sum of intensities of all pixels, which are above and to the left of this specific pixel. Integral image can be calculated with one pass over original picture.

Haar-like features (Fig. 1) are simply differences between sums of intensities of rectangular regions. These features resemble Haar wavelets, hence the name. Usually, three types of Haar-like features are employed: two-rectangular, three-rectangular and four-rectangular.

Sum of intensities of black areas is subtracted from sum of intensities of white areas. Integral image representation allows to calculate Haar-like features in constant time. Still, single detection window contains hundreds of thousands such features, thus it is necessary to select only those that best classify faces from non-faces.

Classifiers are trained using modified version of "AdaBoost" [5] algorithm. Each round of boosting selects single feature that best separates positive and negative samples. Features are added to classifier until a desired detection and false positives rate is acquired.

Each classifier is weak, which means it is only slightly better way of detecting faces then a random guess. It also has low false positives rate. Several such weak

classifiers are joint together to form the attentional cascade in form of a degenerate decision tree (Fig. 2).

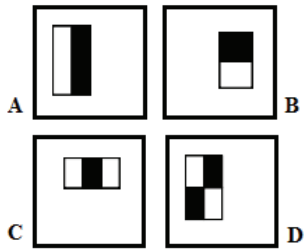

Fig. 1. Haar-like features: two-rectangle (A and B), three-rectangle (C) and four-rectangle (D)

Since each classifier in this cascade has low false positives rate, it rejects a lot of windows in early stages of processing. First stages of cascade are very primitive, thus little processor time is needed to reject unpromising windows and more time can be spent processing windows that might actually contain faces (hence the name "attentional cascade").

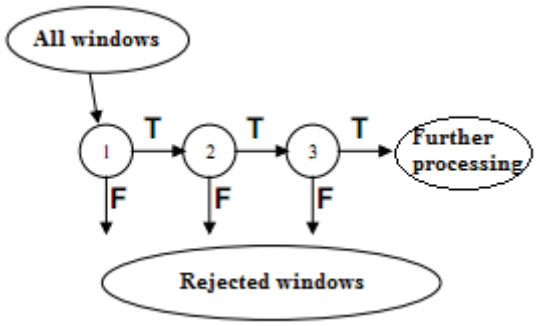

Fig. 2. Attentional cascade

Integral image, modified "AdaBoost" algorithm and attentional cascade comprise the essence of Haar training. It allows for very rapid detection of mostly rigid objects (in our case – faces) with sufficient accuracy.

In our experiments, Haar training algorithms that are implemented in "OpenCV" computer vision library were used. Various classifiers supplied with the library were tested and it was discovered that "haarcascade_frontalface_alt_tree" yields best results. Also we found out in previous experiments that image histogram equalisation (contrast enhancement) does not improve classifier performance. It contradicts suggestions made in literature [6].

Approach

Our inspection robot computer vision system encompasses both of the aforementioned methods. Every image received from camera is analyzed using HoG algorithm. Then, top half of this image is analyzed using Haar training. Only top half of this image was scanned because it is very unlikely to find a face at lower height than average human's waist. This greatly reduces the number of windows needed to process and also, as discovered in our experiments, reduces false positives rate if floor is checked, for instance.

Afterwards it is necessary to combine results from HoG and Haar training. Simple solution can be used: if a face square is within top half of silhouette rectangle, it is considered they both belong to the same person (Fig. 3).

Fig. 3. Combining the results. Face square of the girl in the middle is inside top half of silhouette rectangle, thus both rectangles belong to the same person

Experiment

Standard web camera was used in our experiments. About an hour of footage was taken in corridors of Linköping University with heavy traffic of students. Camera was mounted at around stomach height so people at the end of a corridor were located in the middle of the frame height and were facing the camera without an angle.

Later in the lab, program was written that takes one hundred of random 640 * 480 pixels frames from the footage, makes them greyscale and saves to disk. These frames were processed employing our approach.

Locations of people in every of these one hundred frames were manually marked so method performance can be tested. It was considered that the image to contains a person if at least half of them was visible and not occluded.

Then the algorithms were run on these images with different scale factors. Scale factor determines how big the jump between image scales is. Greater it is, more likely it is to miss certain size faces.

For each scale factor in every frame correct detections (with HoG alone, with Haar training alone and with combination of methods) and false positives (again, with three methods) were manually counted.

Finally, average frame processing speeds of the methods with different scale factors were calculated. Computer with 2.21 GHz microprocessor was used for this task.

Results

Fig. 4 shows how scale factor affects accuracy of HoG alone, Haar training alone and combinations of these methods.

It can be observed that HoG method is way more precise then Haar training in detecting people and it makes perfect sense. A person has to be facing the camera and

close enough to it to be detected which happens rarely. Still, in those rare occasions HoG is worthless because only part of human silhouette is visible when a person is close to the camera (Fig. 5). These are times when Haar training comes in handy.

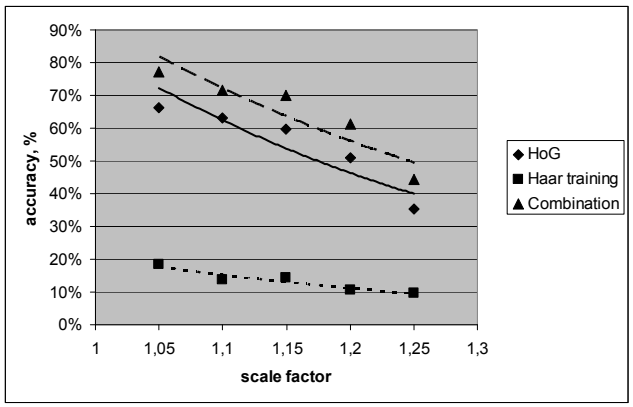

Fig. 4. Accuracy of tested methods as a function of scale factor

Fig. 5. Only part of guy's silhouette is visible, thus HoG classifiers becomes useless

Not only has it proved useful in cases where only part of a human is visible, but it also supplies inspection system with additional information such as whether a person is facing the camera, allows to recognize the face.

With scale factor of 1.05 our approach reaches around 80% accuracy and scale factor of 1.25 still yields reasonable results – 50%. Having in mind that inspection system could process several frames per second with slightly different positions of people, probability that every person will be detected at least once is pretty high.

Fig. 6 shows how scale factor affects false positives rate of HoG alone, Haar training alone and combination of these methods.

Not a single Haar training false positive was experienced in our tests, which makes us conclude it is minute. This also makes false positive rates of HoG method and combination of methods equal. Still, they are pretty steep, especially using higher scale factors.

Finally, algorithm speed results which where acquired using our computer with 2.21 GHz microprocessor are presented in Fig. 7.

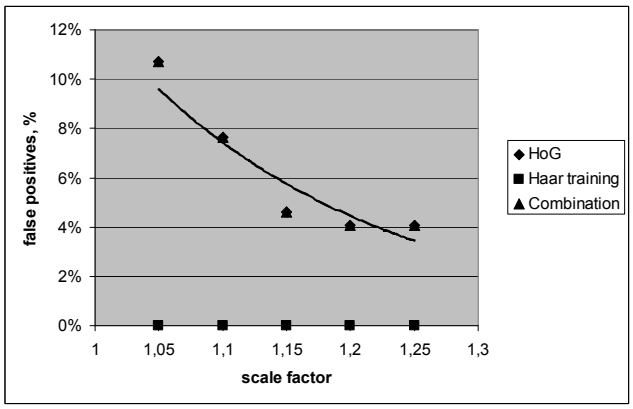

Fig. 6. False positives of tested methods as a function of scale factor

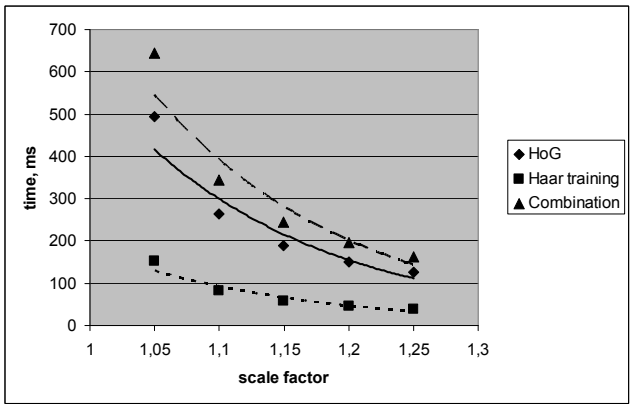

Fig. 7. Average time taken to process a single 640 * 480 pixels frame as a function of scale factor

Using our method a single frame can be processed in around a half of a second with 1.05 scale factor. This is very reasonable, having in mind number of windows analyzed within each frame. If ~50% accuracy and ~4% false positives rate is enough, up to six 640 * 480 pixel frames can be processed in a second.

Conclusions

Inspection system using combination of HoG and Haar training could be used for surveillance purposes. Even though Haar training does not perform well in detecting humans, it complements HoG classifier in cases where part of human silhouette is occluded. Moreover, it supplies inspection system with additional information, allows to tell whether the observed person is facing the camera. Additional methods [7] could be used to further improve performance.

Future work

Since a lot of modern microprocessors have multiple cores, it would be wise to parallelize the algorithm. One of simpler ideas would be to perform HoG calculations on one core and Haar training detection on the other concurrently. Results could then be concatenated. HoG is slower then Haar training, thus lower scale factor could be used for Haar training without slowing down overall processing.

Ability to recognize back of the head would be useful when only part if a person is visible and they are not facing the camera. This could be implemented using additional HoG classifier or some other method.

It is necessary to train stronger HoG and Haar classifiers with bigger datasets. This would improve both accuracy and decrease false positives rate in our system.

It is possible to reduce false positives rate and increase accuracy by using stereo vision [8]. Human silhouette can only be found on the ground and faces – around 1.5 metres above it. Also, a person will always be closer to the camera then its surroundings. Stereo vision could help check whether these conditions are met.

Acknowledgements

This work was partly supported by the Lithuanian Science Council Student Research Fellowship Award (M. P., A. V.)

References

1. **Dervinis D.** Extracting Characteristic Face Points From 2D Image // Electronics and Electrical Engineering. – Kaunas: Technologija, 2004. – No. 5(54). – P. 18–22.
2. **Dalal N., Triggs B.** Histograms of oriented gradients for human detection // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – San Diego, 2005. – No. 1. – P. 886–893.
3. **Burges C. J. C.** A Tutorial on Support Vector Machines for Pattern Recognition // Data Mining and Knowledge Discovery. – Hingham: Kluwer Academic Publishers, 1998. – No. 2. – P. 121–167.
4. **Viola P., Jones M.** Rapid object detection using boosted cascade of simple features // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – Kauai, 2001. – No. 1. – P. 511–518.
5. **Freund Y., Schapire R. E.** A decision–theoretic generalization of on–line learning and an application to boosting // Second European Conference on Computational Learning Theory. – London: Springer–Verlag, 1995. – P. 23–37.
6. **Bradski G., Kaehler A.** Learning OpenCV Computer Vision with the OpenCV library. – Sebastpool: O'Reilly Media, 2008. – 512 p.
7. **Adwan S., Arof H.** A New Approach for an Efficient DTW in Face Detection through Eyes Localization // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 2(108). – P. 103–108.
8. **Lipnickas A., Knyš A.** A Stereovision System for 3–D Perception // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 3(91). – P. 99–102.

Received 2011 02 15
Accepted after revision 2011 06 29

M. Petkevicius, A. Vegys, T. Proscevicius, A. Lipnickas. Inspection System based on Computer Vision // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 10(116). – P. 81–84.

State-of-the-art inspection systems are often used for surveillance purposes, which inspect human behaviour or detect suspicious activities using visual data. Crowd detection and individual person tracking in real-time require robust human detection. We propose a hybrid method, which is based on a combination of histogram of oriented gradients (HoG) and Haar-like features. The former is used for human silhouette detection and the latter for face detection. Novel method ensures more precise human detection and the ability to estimate looking direction. Our method shows better results then HoG or Haar-like features independently. Ill. 7, bibl. 8 (in English; abstracts in English and Lithuanian).

M. Petkevičius, A. Vegys, T. Proscevičius, A. Lipnickas. Kompiuterine rega paremta inspekcinė sistema // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 10(116). – P. 81–84.

Modernios inspekcinės sistemos dažnai naudojamos priežiūros reikmėms. Jos stebi žmogaus elgseną ar bando aptikti įtartiną veiklą, naudodamosi vaizdine informacija. Minios aptikimas bei atskirų žmonių sekimas turi būti patikimas. Mes siūlome hibridinį metodą, paremtą gradiento krypčių histogramomis (HoG) bei Haaro požymiais. Pirmasis naudojamas aptikti žmogaus siluetui, o antrasis – žmogaus veidui. Naujasis metodas leidžia tiksliau aptikti žmones bei nustatyti, kur žiūri žmogus. Mūsų metodu gaunami geresni rezultatai nei naudojant HoG ar Haaro požymius atskirai. Il. 7, bibl. 8 (anglų kalba; santraukos anglų ir lietuvių k.).