PEOPLE REIDENTIFCATION IN A DISTRIBUTED CAMERA NETWORK

Icaro Oliveira de Oliveira, Jose Luiz de Souza Pio

Federal University of Amazonas - Computer Science Department. E-mail: ioo@dcc.ufam.edu.br/ josepio@dcc.ufam.edu.br

RESUMO

Este artigo aborda o problema da reidentificação da presença de pessoas em um ambiente de câmeras interconectadas por um sistema de comunicação. O problema da reidentificação ou reaquisição consiste no processo de realização dinâmica da correspondência entre imagens extraídas por câmeras diferentes. A metodologia aqui desenvolvida utiliza técnicas de detecção de pontos de interesse e cor dos alvos por meio do espaço *wavelet*. A aplicação deste trabalho é importante para os modernos sistemas de segurança nos quais a identificação da presença de alvos no ambiente monitorado amplia a capacidade de atuação dos agentes de segurança em tempo real e fornece a localização de cada pessoa de forma imediata. Resultados satisfatórios foram alcançados através dos experimentos reais por meio dos pontos de interesse aplicados em grandes bases públicas de vídeos e imagens sintéticas com ruído.

Palavras-chave: reidentificação, wavelet, rede de câmeras.

ABSTRACT

This paper presents an approach to the object reidentification problem in a distributed camera network system. The reidentification or reacquisition problem consists essentially on the matching process of images acquired from different cameras. This work is applied in a monitored environment by cameras. This application is important to modern security systems, in which the targets presence identification in the environment expands the capacity of action by security agents in real time and provides important parameters like localization for each target. We used target's interest points and target's color with features for reidentification. The satisfactory results were obtained from real experiments in public video datasets and synthetic images with noise.

Keywords: reidentification, camera network, interest points.

1 INTRODUCTION

In many video-surveillance applications, it is desirable to determine if a presently visible person has already been observed somewhere else in the cameras system. This kind of problem is commonly known as ”object reidentification”, and a general presentation of this field can be found for instance in (TU et al., 2007). Re-identification algorithms have to be robust even in challenging situations caused by differences in camera viewpoints and orientations, varying lighting conditions, pose variability of persons, and rapid change in clothes appearance.

The proliferation of video cameras in public places such as airports, streets, parking lots and shopping malls can create many opportunities for business and public safety applications, including surveillance for threat detection, monitoring parking lots and streets, customer tracking in a store, assets movement control, detecting unusual events in a hospital, monitoring elderly people at home, optimization of employee placement etc. These applications require the ability of automatically detecting, recognizing and tracking people and other objects by analyzing the video or image data. In spite that video surveillance has been in use for decades, the development of systems that can automatically detect and track people is still an active research area.

Many approaches have been proposed in recent years. They differ in various aspects such as the number of cameras used, the type of cameras (grayscale or color, mono or stereo, CCD or Webcams etc) and their speed and resolution, type of environment (indoors or outdoors), area covered (a room or a hall, a hallway, several connected rooms, a parking lot, a highway etc), and location of cameras (with or without overlapping fields of view). There is a vast number of papers that are devoted to this research area and we shall overview some of them in the related work section below. However, the performance of most systems is still far from the requirement of real applications due to three major reasons. First, actually the capability of the video and image analysis techniques are dwarfed by the complexity of the problem. For example, The existing face recognition techniques for person identification in an image, are susceptible to many factors, such as changes in the luminosity, view angle and even very simple disguises. Second, most of such systems are developed and tested in controlled ”simulated” environments, e.g. a person deliberately walking back and forth in the room, instead of the real public areas where events take place naturally. Finally, the efforts have been mostly focused on analyzing the video data, without taking into consideration the domain knowledge about the environment that can essentially improve the accuracy of both tracking and recognition.

The amount of surveillance videos is increasing. Finding useful information from a large collection of surveillance videos is becoming unmanageable by human. This paper aims at building an object retrieval system for such a large collection. The detection and matching of human appearance is a challenging problem. The viewpoint changes lead to the difference in appearance. The lighting conditions are not constant due to sunlight, street light, and cloud. Occlusion happens frequently due to the presence of other objects and traffic. Weather condition such as rain also causes severe changes in background as well as foreground.

2 RELATED WORK

Several studies have contributed to this theme, such as tracking methods presented in Wei et al. (2004), Hamdoun et al. (2008), Morioka et al. (2006) and Park et al. (2006), methods of persons detection as used in Pham et al. (2007). The novelty presented in this paper is the match of interest points, different from the methods

developed (GHEISSARI et al., 2006) that uses matches among images (WEI et al., 2004) that uses biometric characteristics (PARK et al., 2006; MORIOKA et al., 2006; PHAM et al., 2007) and that uses color matches.

The approaches (HAMDOUN et al., 2008; PARK et al., 2006; CLEMENS et al., 2007) have limitation on the number of people to be reidentified, because when there is movement of people or cars a new target is inserted into the structure as it also occurs (HAMDOUN et al., 2008) that does not treat the movements of people as it is treated (GHEISSARI et al., 2006). The structure used in Clemens et al. (2007) for cars has a method of maintenance for removing unnecessary nodes, but this additional procedure has become a bottleneck in the system. In Hamdoun et al. (2008) a person reidentification scheme using matching of interest points collected in several images during short video sequences was proposed and evaluated, while in this work, a a person re-identification scheme using matching of interest points from one image is proposed and evaluated.

In this work and in these approaches (OLIVEIRA; PIO, 2009a, 2009b, 2009c) we present a novel algorithm for object reidentification in camera networks. The major contributions are the very efficient description of targets using local features, and the subsequent communication and matching algorithm proposed. First, the operator sent a image with the interest target. This image is identified and described by robust local features. The object description is subsequently converted into an extremely compact representation, so called signature. This signature might further be communicated to the neighboring camera nodes. The objects seen in a camera node are again described and a specific signature for the object is created. The point correlation algorithm in each camera node compares the camera signatures with the operator signature and allows for efficient reidentification and tracking of the target through entire camera networks. Our approach different the approaches (HAMDOUN et al., 2008; PARK et al., 2006; CLEMENS et al., 2007) requires no topological knowledge of the environment, training standards of objects or data structure. Additionally, one major goal of our work was to use uncalibrated setups since multi-camera calibration is still a tedious task. The power of our approach is shown on the simulation of several traffic scenarios; first, a one-camera setup is considered in several scenarios and then a two-camera network is simulated; second, a scenario using synthetic images with several Gaussian noises. The results showed our approach was useful and motivate further research in the area of local features for object reidentification.

3 OBJECT REIDENTIfiCATION

As former mentioned, the entire object reidentification approach presented in this paper is solely based on local appearance features. Although the usage of local features has some shortcomings, i.e. computational costs, minimum requirements on image quality, etc., and highly relies on robust matching methods, it has the advantage of being partially robust to image scale, rotation, affine distortion, addition of noise and illumination changes. By using a setup such as the proposed one, it is not necessary to tediously train classifiers for each camera or add information of additional cues or sensors. In the following, we shortly introduce our choice of local features and describe the point correlation algorithm.

3.1 Interest Point Detection

Interest point detection reduces the complexity of visual recognition, by determining the salient points in an image. A ”point” actually refers to a region of the image with a regular

shape, centered on a specific point. This regular shape may circular (or elliptical, if affine invariance is desired), or it may also be rectangular (a cruder approach that is more straightforward for calculations). Analyzing a region surrounding a point is necessary because a single pixel does not have enough information to be deemed interesting or uninteresting.

Only the points found with the interest point detector will be passed to the feature descriptor. This does not require, however, that every pixel fall into at least one interest point's area. Certain regions of the image may not be useful in object recognition, and thus should be ignored. Ideally, portions of the image that are ignored by the interest point detector do not contain useful information for object recognition.

It would seem that if a large percentage of the image is covered by the detected interest points, then little has been accomplished. The argument would be that only a small percentage of the information has been eliminated. This is not the case. A good detector will properly aggregate the pixels of the image, so that they are easily analyzed in the feature description phase.

Lowe (2003) approximated the Laplacian-of-Gaussian using a difference-of-Gaussian function. This function performs two consecutive smoothings using a Gaussian, and finds the difference in the resulting images. This is efficient to compute because the smoothing is already necessary for building multiple scales in the image pyramid.

In this paper, we use the Bay et al. (2006) detector. This detector is a detection scheme that is faster than the Lowe. It relies on integral images (VIOLA; JONES, 2001) and box filters to approximate the determinant of the Hessian matrix[1]. This detector is more distinctiveness, and robustness, yet can be computed and compared much faster. This detector have high repeatability.

3.2 Feature Description

In this stage, the interest points detected are descripted in color and SURF Description.

3.2.1 Color

This section describes how the feature vector by color is generated for given interest point. The first component of this feature vector is a histogram of the hue of the point. To overcome the sensitivity of HSV histograms to changes in illumination and shadows in outdoor scenes, we use the definition of hue which is invariant to brightness and Gamma (SWAIN; BALLARD, 1990). This definition of hue shown (SWAIN; BALLARD, 1990) for a given RGB color space is as follows:

$$H = arccos \frac{log(R) - log(G)}{log(R) + log(G) - 2 * log(B)} \quad (3.1)$$

This color is utilized in the signature generation process.

3.2.2 SURF Description

A feature descriptor attempts to characterize a region in a robust way that is invariant to natural viewing changes. This may include scale, rotation, lighting, and affine/viewpoint variance. The SURF (BAY et al., 2006) descriptor acknowledges the high matching performance of SIFT (LOWE, 2003) and attempts to replicate this performance while improving speed. The region of interest is divided into a square grid of 4 ×4 sub-regions, like SIFT (LOWE, 2003). However, instead of using a histogram of image gradients to characterize each sub-region, SURF measures the response of Haar wavelets summed over the subregion. Both a

[1] Hessian Matrix is the square matrix of second-order partial derivatives of a function. This matrix describes the local curvature of a function of many variables

"vertical" and "horizontal" Haar wavelet [2] response are calculated. The actual directions are relative to the interest point's orientation. There are four values in the descriptor for each region: the sum of the response for each wavelet, and the sum of the absolute value of the response for each wavelet. This SURF description is utilized in the signature generation process.

3.3 Signature Generation

Recently encouraging recognition results have been achieved in the area of large-scale image databases, using tree-like representations of local descriptors. One approach is the usage of vocabulary trees and inverse voting, like the work of Clemens et al. (2007) and Hamdoun et al. (2008). This approach don't use vocabulary tree, instead our framework utilize the advantages of efficiently build a compact object signature through query image provided by operator. First, the operator provides data through an image with the interest target, as shown in Figure 1. This image is obtained for motion detection or tracking in the camera selected by the operator.



Figure 1. The image provided by the operator.

In this image, we detect the *n* interest points with Fast-Hessian detector. After, we extract the SURF descriptor vector or/and the color to each interest point detected. This vector or/and this color is stored in a compact signature. This signature is sent to all cameras in system. In each camera the point correlation is made between the interest target signature and the descriptor vector of the interest points detected in the frame, shown in Figure 2, of each camera in the system.



Figure 2. The frame provided by the camera.

3.4 Points Correlation

This method doesn't utilize vocabulary tree, instead a signature of the interest target and a signature of the camera *i* frame per second. This method is efficient, because the signatures are relatively short vectors (typically in the order of a few hundred elements) and can be matched very fast, sorting of the elements in increasing order still simplifies the algorithm complexity and makes more efficient matching possible. The points correlation process is formed by logical and mathematicals operations.

The vector x is formed by *m* interest points obtained interest target signature and the vector ci is formed by *p* interest points obtained the camera *i* frame. The points correlation is a matching process between *m* interest points x_j, $j = 1,...,m$ and *p* interest points $c_{i,j}$, $j = 1,..., p$. Each interest point has a feature vector.

We find the Laplacian signal for each interest point in x_j and $c_{i,j}$. This signal represents the maximum positive signal and the maximum negative signal in Hessian matrix. For each point descriptor xj and ci,j with the same signal the sum of quadratic differences (SQD) is applied

[2] The Haar wavelet is a certain sequence of functions. This sequence was proposed in (ALFRED, 1910)

$$\sum_{y=1}^{d} (k_y - l_y)^2, \qquad (3.2)$$

where k is an interest point vector descriptor in xj, l is a interest point vector descriptor in ci,j and d is amount of dimensions the vector. If the vector is formed by SURF description then d is 64 or 128 else if the vector is formed by color then d is 1 else if the vector is formed by color and by SURF description then d is 65 or 129. A matching pair is detected if its SQD distance is closer than L times the distance of the second neighbor, where L is a threshold. This means that the set of interest points in x matches ci. In other words, the interest target is in the camera *i* frame. Finally, this frame and *i* are communicated to the operator. Thus, the operator receives all frames of the cameras system correlated interest target.

4 SYSTEM EVALUATION USING REAL VIDEOS

This section presents two groups of experiments. The first group uses a synthetic image database to assess the robustness of the combination of the recognition method against the feature vectors using the noise insertion and scalability of people. The second group uses public video datasets with the objective of forming videos databases with heterogeneous informations like environments, the camera view angles, number of people, cameras with overlapping and non-overlapping fields of view, different types of cameras and movements. This group receives most robust features vector obtained from group 1. This group evaluated accuracy features vector in the person reidentification on real video and the points correlation complexity is shown. These evaluations groups are described in the following sections.

4.1 Specification of the hardware used

The settings used in the experiments are Configuration 1 with Pentium 4 Core Duo 3.2GHZ, 2GB RAM, 250 GB, C++ Language, OPENCV and OPENSUSE 11.1. Configurantion 2 is a notebook amazon pc tablet l97, C++ Language, OPENCV and OPENSUSE 11.1.

4.2 Group 1: System evaluation using synthetic images

This group analyzes the system's robustness on the scalability of 38 interest objects. This group disregarded a camera system, because the specific objective of this evaluation is to check the robustness of the features recognition of a query on a frame with several interest objects. For each object (3D model person), we check the robustness of the features recognition against the Gaussian noise insertion in the query image and in each object in the frame.

In this evaluation group, the matching method analyzed the three feature vectors with the threshold 0.7 and 0.8. The three feature vectors are color, interest points, color and interest points. The interest points were described in 128 dimensions. Consider the size of the noise ranged from 91 to 121 and the scalability of models made by 26 people and 12 others models. The objective is to evaluate the robustness of the combinations between the matching method and the three feature vectors. Consider also the scale insertion 1/1.5 X, 1/3.0 X and ¼.5 X in the query images. In the follow sections the synthetic image database and the evaluation details will be shown. A histogram equalization process is performed to increase the image sharpness about query images with scale. After, the noise is added in the image.

4.2.1 Syntethic images database

This database is formed by 26 images of 3D models people and 12 images of 3D models objects in a total of 38 different models from Google Sketchup software. The synthetic images are shown in Figure 3.



Figure 3. On the left side images of 3D models people and, on the right side, images of 3D models objects.

4.2.2 System evaluation with the Gaussian noise insertion in the query image and in each object in the image

In this evaluation, we inserted Gaussian noise with the size ranged from 91 to 121 in 3D models people used as query and in all 3D models used as object in the image, where the recognition robustness is checked. This size is high, because in the experiments the *size* < 91 is stable. The details of the results are presented below. Each graph of the results is the average quantity of the hits between the results of each scale (1/1.5 X, 1/3.0 X and 1/4.5 X) inserted in the image query.

Feature vector formed by interest points

In Figure 4 when the noise increases in the query image, the accuracy decreases. While between the noises 95 × 95 at 97 ×97, 99 ×99 at 101 ×101, 103 ×103 at 107 ×107, 109 ×109 at 111 × 111, 119 × 119 at 121 × 121, the accuracy is stable, because the features have stable results in this noise levels.

In most cases the threshold 0.8 has shown higher accuracy on high noises than the 0.7 threshold, because the threshold 0.7 remove some correct images. This matching method with the interest points shown to be robust with an accuracy between 5% and 19% on high noises levels.

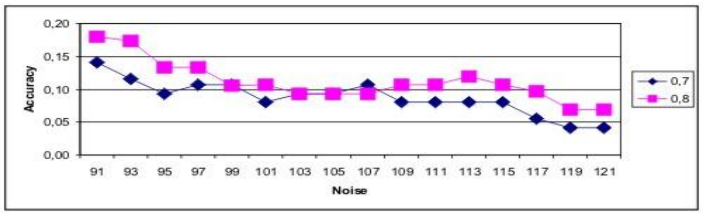


Figure 4. Matching results with the Laplacian signal and the feature vector formed by the interest points.

Feature vector formed by color

This feature vector is obtained in 3.2.1. In Figure 5 when the noise increases in the query image, the accuracy decreases, because the image noise is the random variation of brightness or color information in images. In other hand, the image becomes dithering. So, the color information is variant in differents noise levels. However, it is perceived that there are several moments in which the feature is stable, in other words, the information color is invariant at the noise. In these cases the accuracy is constant. But it is observed that the accuracy is 4% in the best case. This is unsatisfactory because it is below the worst case of the previous experiment. This happens because the color is a feature variant at light, at noise insertion and principally of scale. On the other hand, the interest point is not formed by the color.

Feature vector formed by color and interest points

In Figure 6, the threshold 0.7 becomes unrecognizable in some noises. In this case the color used in the feature vector influenced the matching to find false positives in all queries. Therefore, this experiment is poor.

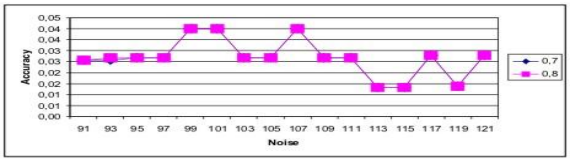


Figure 5. Matching results with the Laplacian signal and the feature vector formed by channel

H.

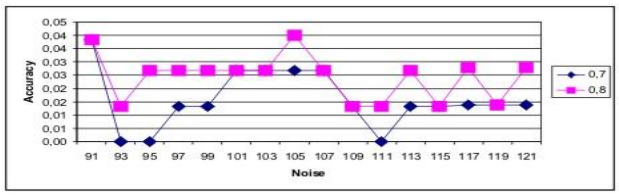


Figure 6. Matching results with the signal Laplacian and the features vector formed by the channel H and the interest points.

4.3 Group 2: System evaluation using videos

The group 2 of experiments evaluates the reidentification accuracy on real videos. This group evaluates the query images accuracy provided by the operator on real videos. This query image contain only one view angle person. In other words, the approach reidentificated with this query image multiples view angles of this person with rates over 82% accuracy, which is different from the work developed in Hamdoun et al. (2008) that uses a model of each person storing all angles for reidentification.

4.3.1 Video databases description

In Cantata (2000) there is a large amount of public video datasets about people, cars and medical images. In this database were obtained, as:

- base1: Obtained in (ACTIONS, 2005). This database has high irregularities in the video sequences. It has 10 people videos walking in different scenarios with different backgrounds not uniform and performing several movements. These videos have low sensitivity to partial occlusions and deformations in each non-rigid motion. The database has pre-classified 7,133 in movements and in each person. In the images acquisition process was set to 1 fps, since the amount of the database is small.
- base2: Obtained in (CAVIAR, 2004). This database provides an amount of videos recorded in different scenarios with people walking into a mall in a shopping centre in Lisbon. For each sequence, there are two time synchronised videos, one with the view across and the other along the hallway. The resolution is half-resolution PAL standard (384 x 288 pixels, 25 frames per second) and compressed using MPEG2. Unlike the base1, this database is not pre-classified in people and it has 72,466 images. In the images acquisition process was set to 25 fps, since the amount of database is large.

4.3.2 Query database description

The query images were obtained from public videos databases. Each image represents a view angle of each person, which is different from (HAMDOUN et al., 2008) that uses a model of each person with different view angles of the same as query. These images are obtained by a motion detection or tracking. Figures 7 and 8 show the query of each database.



Figure 7. These query images represent base1 people, which are extracted from person frontal view with little deformation.



Figure 8. These query images represent base2 people, which are extracted from person frontal view and back view with little deformation and with panoramic view.

4.3.3 Reidentification Evaluation

It considers feature vector formed by interest points (64 and 128 dimensions) and the matching method using Laplacian signal. In this paper the reidentification considers at least corresponding 2

points between video frame and query image. This article was examined on four thresholds 0.5, 0.6, 0.7 and 0.8.

This evaluation verified the accuracy of the features vector formed by interest points on the public video databases using the query images for each database. The result is shown in Figure 9.

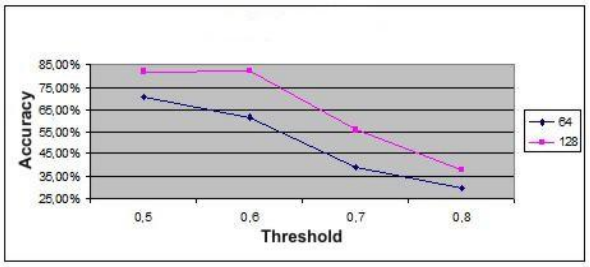


Figure 9. Feature vectors accuracy on video database.

Figure 9 shows that the function is decreasing on the threshold 0.5 in the vector of 64 dimensions. This shows that increasing the threshold, the error rate also increases leading to decrease in accuracy. However, the vector of 128 dimensions, the function is decreasing after the threshold 0.6. This shows that the increase in error rate occurs when the threshold is greater than 0.6. It can also view that as the threshold increases the 2 functions (vector of 64 dimensions and the vector of 128 dimensions) will always be parallel. The system achieve 82,2% with best result.

For comparison, the best result reported (HAMDOUN et al., 2008) on a set of videos including 10 different people is 82% correct best match. Ours contain two set of videos including 15 different people in several scenarios with overlapping and non-overlapping cameras. In this paper, we achieve maybe 82,2%. It shows difficult examples of reidentification below, because they show changes in their clothes and in some cases little occlusion. The query reidentification is using the interest points vector with 128 dimensions and a threshold with 0.6.



Figure 10. A query reidentification shows people near the person who it wants to reidentify.

Figure 10 shows differences in clothes and the person's placement. However, the system has reidentified the interest points of the person's face and person's shirt.

Figure 11 shows the person in the image query with occlusion in the camera frame rei-deintified. The operator could have difficult in finding this person in each video, but the system manages to identify this case from the interest points of the person's face in the image query.

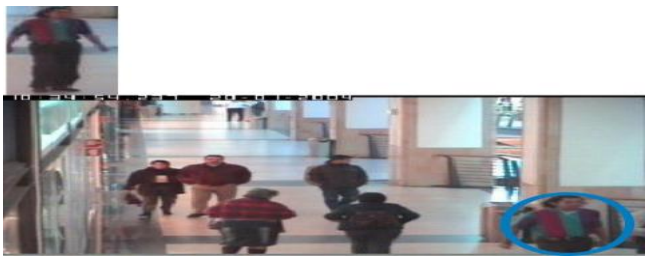


Figure 11. A query reidentification shows people near the person who it wants to reidentify and this person has occlusion.

In each testing, the approach localizes the query object through query object's interest points on one object or multiple objects obtained in the camera frame.

4.3.4 Points correlation Complexity

The points correlation complexity is formed by the number of q camera frame points and the number of points c of the query. Thus, the points correlation complexity is $O(q * c)$. The work developed (HAMDOUN et al., 2008) uses in the

recognition a kd-tree with complexity $O(n \log n)$. The number of points c obtained from one query image is less than the n points obtained in 21 images used with query in (HAMDOUN et al., 2008). Thus, this points correlation method is more efficient than the recognition algorithm developed in (HAMDOUN et al., 2008).

5 CONCLUSIONS AND FUTURE WORK

In this paper we presented a novel approach for object reidentification in the cameras system. The proposed algorithms are designed to run fully autonomous on network of cameras and are real-time capable. While communication overhead between adjacent camera nodes is notably reduced, we proof the application of high-level local features advantageous for object signature. While the application of the approach proposed is not limited to a special object category, our algorithms were tested on traffic scenarios and encouraging results were presented. The main limitations of this article is the necessity of experiments with real cameras network and the cameras have a local processing. Although the proposed method in the task of helping people identification based on the descriptor points of interest, other evidence can be proposed through the insertion of new attributes, such as people identification through selected points or regions in images. Extensions of the proposed article can be a search engine for videos of people online, the 3D models reidentification or even other interest objects such as cars and lost objects. This system can integrate with others recognition or tracking systems.The deployment of the system in real environment is an activity that should be complemented in the future.

REFERENCES

ACTIONS. **Actions as space-time shapes**. 2005. Online in: http://www.wisdom.weizmann.ac.il/ vision/SpaceTimeActions.html. Accessed in: 10 Aug. 2008.

ALFRED, H. Zur theorie der orthogonalen funktionensysteme. **(German) Mathematische Annalen**, n.69, p.331-371, 1910.

BAY, H.; TUYTELAARS, T.; VAN GOOL, L. SURF: Speeded Up Robust Features. In: proceedings of the Ninth European Conference on Computer Vision. **Proceedings…** 2006.

CANTATA. **Datasets for CANTATA project**. 2000. Online in: <http://www.hitech-projects.com/euprojects/cantata/datasets cantata/dataset.html>. Accessed in: 10 Aug. 2008.

CAVIAR. **CAVIAR test case scenarios**. 2004. Online in: <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>. Accessed in: 10 Aug. 2008.

CLEMENS, A.; CHRISTIAN, L.; HORST, B. Object reacquisition and tracking in large-scale smart camera networks. In: ACM/IEEE International Conference On Distributed Smart Cameras, 1. **Annals…** Viena, 2007.

GHEISSARI, N.; SEBASTIAN, T. B.; HARTLEY, R. Person reidentification using spatiotemporal appearance. In: Conference On Computer Vision And Pattern Recognition. **Annals…** 2006. p.1528–1535.

HAMDOUN, O.; MOUTARDE, F.; STANCIULESCU, B.; STEUX, B. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In: ACM/IEEE International Conference On Distributed Smart Cameras, 2. **Annals…** 2008. p.1–6.

LOWE, D. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision,** v.20, p.91–110, 2003.

MORIOKA, K.; MAO, X.; HASHIMOTO, H. Global color model based object matching in the multi-camera environment. In: Intelligent Robots and Systems, 2006 IEEE/RSJ INTERNATIONAL CONFERENCE. **Annals…** 2006, p.2644–2649.

OLIVEIRA, I.; PIO, J. Object reidentification in distributed cameras system. In: IEEE International Symposium on Ubisafe Computing, 2. **Annals…** 2009.

OLIVEIRA, I.; PIO, J. Object reidentification in multiple cameras system**.** In: International Conference on Embedded and Multimedia (EM)

COMPUTING, 4. **Annals...** 2009.

OLIVEIRA, I.; PIO, J. People reidentification in a camera network. In: International Workshop on Multimedia, Information Privacy and Intelligent Computing Systems, 2. **Annals…** 2009.

PARK, U.; JAIN, A.; KITAHARA, I.; KOGURE, K.; HAGITA, N. Vise: Visual search engine using multiple networked cameras. In: Pattern Recognition, International Conference on, IEEE Computer Society. **Annals…** Los Alamitos, CA, USA, v.3, p.1204–1207, 2006.

PHAM, T.; WORRING, M.; SMEULDERS, A. **A multi-camera visual surveillance system for tracking of reoccurrences of people**. 2007. Online in: <http://dare.uva.nl/cgi/arno/oai/uvapub>.

SWAIN, M. J.; BALLARD, D. H. Indexing via color histograms. In: Proceedings of the Darpa Image Understanding Workshop. **Proceedings…** Pittsburgh, PA, USA: [s.n.], 1990.

TU, P.H.; DORETTO, G.; KRAHNSTOEVER, N.O.; PERERA, A.G.A.; WHEELER, F.W.; LIU, X.; RITTSCHER, J.; SEBASTIAN, T.B; YU, T.; HARDING, K.G. **An intelligent video framework for homeland protection**. 2007. Online in: <http://vision.csee.wvu.edu/publications/tuDKPWLRSYH07SPIEdss.pdf>

VIOLA, P.; JONES, M. Rapid Object Detection Using a Boosted Cascade of Simple Features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **Proceedings…** 2001. p.511–518.

WEI, G.; PETRUSHIN, V. A.; GERSHMAN, A. V. Multiple-camera people localization in a cluttered environment. In: Proceedings Multimedia Data Mining Workshop. **Proceedings…** 2004.