Human Body Detection in the RoboCup Rescue Scenario

Shahram Bahadori, Luca Iocchi
Dipartimento di Informatica e Sistemistica
Università degli Studi di Roma "La Sapienza"
Via Salaria 113, 00198 Roma Italy
E-mail: <lastname>@dis.uniroma1.it

***Abstract*— This paper presents an analysis of techniques that have been studied in the recent years for human body detection (HBD) via visual information. The focus of this work is on developing image processing routines for autonomous robots operating for detecting victims in rescue environments. The paper both discusses problems arising in human body detection from visual information and describes the methods that are more adequate to be applied in a rescue scenario. Finally, some preliminary experiments for such methods in recognizing rescue victims are reported.**

I. INTRODUCTION

One of the main objectives of a search and rescue mission is to rescue the disaster victims. The rescue environment may be very hazardous for the rescue human agents, because of the unstable situation of the camp, therefore robotic rescue agents can be important collaborators for people involved in a rescue mission. One of the most important requirements in such rescue scenarios is to identify the victims from all the other objects. This task is very simple for a human agent, while it is not for a robotic agent. However, there are situations in which human agents cannot (safely and quickly) intervene and thus the ability of detecting human victims is important also for rescue robots that can access such dangerous areas.

In order to detect a human body, an autonomous robot must be equipped with a set of specific sensors, that provide information about the presence of a person in the environment around. Each of these sensors is suitable to detect a specific characteristic of human bodies, such as shape, color, motion, IR signals, temperature, voice signals, CO_2 emissions, etc. Although sensor fusion is a fundamental issue in this context and an effective implementation of a human body detection system should integrate information coming from several kinds of sensors, in this paper we focus our attention on the problem of human body detection from visual information. The main reason is that, among all the sensors that may be used for detecting human bodies, most of the research work in the literature make use of vision systems, due to the fact that cameras are commonly used in robotics and of course in computer vision and images are rich sources of information, that are usually used also for other tasks, like localization, mapping, obstacle avoidance, etc.

In fact, human body detection by visual information has been studied also in different application fields, that are for example autonomous vehicle navigation for detecting the people and pedestrians for autonomous navigation [1], surveillance [2], human motion capture for virtual reality [3], etc. A large amount of research on HBD has been carried out in the past, specially in the community of Computer Vision (see [1] for a survey), with the main goals of devising techniques for and implementing systems that from one side are able to detect the existence of a human in a scenario and from the other side are able to track and follow the detected mobile object, but in the next sections we will see that non of this methods are sufficient to be consider an autonomous technique for HBD.

The general aspect of HBD by visual information is to acquire images via visual sensors and then processing these images for achieving to the significant results [4], [5]. The difficulties of such processing stem from the number of degrees of freedom (DOF) in the human body, self-occlusion, appearance variation due to clothing, and the ambiguities in the projection of a 3D human shape onto the image plane. In a rescue scenario it is even harder because of the critical situation of the environment and the possibility of being covered or trapped by the rubbles.

The RoboCup Rescue league has been founded in order to realize simulation and robotic systems for supporting rescue operations [6]. The main objective of the Rescue Robot League is to build mobile robots that can actually navigate rescue test arenas [6] aiming at detecting human victims (represented as mannequins). In the past competition (Fukuoka 2002), all of the prototypes of the rescue robots participating in the tournament, were only partially autonomous. In particular, for HBD all the robots were teleoperated and thus the identification of victims in the test arenas were made by team members by visual or IR information transmitted by the robot [7].

Since, we are mainly interested in research of autonomous rescue robots, it is fundamental to devise effective methods for HBD. In this paper we describe

some techniques for HBD by visual information that have been studied in the past in different application fields and we discuss their use in the RoboCup Rescue scenario by showing some preliminary results of their application. A more comprehensive classification and analysis of vision based techniques for HBD is available in [8].

II. Difficulties in Human body detection by visual information

Automatic human detection and body part localization are important and challenging problems in computer vision. The solution to these problems can be employed in a wide range of applications such as safe robot navigation, visual surveillance, human-computer interface, and performance measurement for athletes and patients with disabilities, virtual reality, figure animation and also for search and rescue missions. However, retrieval by shape is still considered one of the most difficult aspects of content-based search [9], and the problems arising in HBD are even more difficult, since human bodies have articulated parts and deformable shapes. Therefore the research on human detection and body part localization is very active and it has produced a wide range of applications on general object detection and shape analysis. The main problems that make the automatic machine human body detection as one of the still open research fields in Artificial Intelligence can be divided in two groups: *the physical appearance problem* and *the classification and individuation problem*.

Physical aspect problems stem from the number of degrees of freedom (DOF) in the human body, from self-occlusion, and appearance variation due to clothing, race, age and the ambiguities in the projection of a 3D human shape onto the plane image. To locate the joints of a person is even harder, because these are hidden by muscle, skin, and clothing. Furthermore, in a rescue scenario, we have to solve the problem of the body occlusion by the external materials (such as rubbles).

For the classification and individuation problems the issue to be considered is *how to classify a previously segmented object as human or non-human*. This is an object classification problem, that is usually addressed with an object representation model and a classifier (see for example [1]. Object classification by shape is difficult mainly because of the noise effect on the sensors, variation of the view points, flexibility of the objects such as human body, the occlusions of the object and in the case of the human body the variety of the bodies due to race, age, etc.

III. Approaches to Human Body Detection by visual information

In the way of detecting an object in an image frame there are several methods. In particular, human body detection is generally addressed by using a three step method. The first step is to find the margins of each object in an image (*Segmentation*): in this phase we can individuate the existence of generic objects, that are simply represented as set (or clusters) of connected pixels. In the second step such clusters of pixels are distinguished and classified on the basis of some predefined classes, representing the various parts of the human body (*Classification*). Finally, in the third step the classified components are composed in order to match a human body model (*Modeling and recognition*). Consider that the second and third steps can be used in an iterative cycle in order to detect the body with better effectiveness and reliability.

In the following of this section we describe these three steps in more details providing advantages and disadvantages for each class of solutions proposed in the literature. This analysis has been explicitly carried out in order to identify those methods that are more adequate for implementing the vision system of an autonomous rescue robot.

A. Segmentation and contour extraction

Segmentation of images is the process of finding the bounds of each object in the scene by contour detection. Segmentation methods can be categorized into four classes [10]: *edge-based* [11], *clustering-based* [12], [13], *region-based* [14], [15], and *split/merge approaches* [16], [17].

a) Edge based approaches.: Image edges are detected and then linked into contours that represent the boundaries of the object image. A very successful method was proposed by Canny [11], according to which the image is first convolved by the Gaussian derivatives, candidate edge pixel are isolated by the method of non-maximum suppression and then they are grouped by threshold, the main advantages of the edge-based methods is their lower computation. However, the edge grouping present serious difficulties in the setting of appropriate thresholds and producing connected, one-pixel-wide contours [18].

b) Clustering based approaches.: In this class of methods image pixels are sorted in increased order as histogram according to their intensities, several Clustering-based approaches have been proposed such as Fussy-c-means (FCM) [16] and K-means [12]. The main advantage of these approaches is that the problem of setting thresholds can be avoided by interactive process. Moreover the segmented contours are always continues. However, over-segmentation may occur because the pixel in this same cluster may not be adjacent. Usually, merge process are applied to solve this over-segmentation problem.

c) Region-based approaches.: The goal of region-based approaches is to find the regions that satisfy a certain predefined homogeneity threshold. The computation time of these approaches is short. However, different

similarity threshold setting may lead to different segmentation results [17]. Furthermore, both watersheds and pyramidal segmentation may cause over-segmentation.

d) Split/merge approaches.: In these approaches an input image is first tessellated into a set of homogeneous primitive regions. Then, similar neighboring regions are merged according to a certain decision rule. These algorithms do not use any threshold, but most of them suffer from a slow rate of convergence.

In a rescue environment, the segmentation phase is very difficult since the signal is desturbed by the noise and normally the bodies are coverd by the external materials. In order to implement an effective segmentation method for rescue robots it is necessary to take into account the following features: *extracting continuous contours*, *Non-over segmentation*, *Non-thresholds* in addition of the above fundametal characteristic an ideal method have to has an acceptable Computation time.

Among the methods for image segmentation, we believe that rescue robots may take advantage of Clustering based segmentation methods becuse the Contours of this method are almost allways continuous and the over segmentation is resolvable by merging the homogeneous segmented areas after segmentation phase, also in this method we dont need to use any thresholds.

B. Object classification

After the image segmentation phase, it is necessary to identify significant objects in the segmented area. There are several methods that can be used for object classification but only a few of them are precise enough to classify the human body parts [1].

In order to classify an object within a class of homogeneous objects it is important to have a good object class model. Such a model should allow the recognition of objects independently of their positions, orientations, sizes, and articulation for articulated objects. It should also accommodate variations among the instances of an object class and should be insensitive to objects with partially missing parts. Human modeling is an essential part of model-based human detection. Although a great number of human models have been proposed in the literature, only a few of them are appropriate for human detection. Most models are developed for other purposes, such as human tracking [19], [20] or figure animation [21]. These models are either too complicated to be practical for efficient human detection, or can just be used to detect a particular person rather than all instances of humans.

A good human body model for automatic search of human bodies from images has to be invariant to scale, orientation, position variation, and it also must be robust to shape distortion due to digitizing and processing, and have to allow the articulated moving. The large collection of proposals for human modeling [1] has been classified in the following in five categories:

- *Part-based representation*
  There are vary part-based representations to handle articulation. They vary widely in their level of detail. In particular, we distinguish 2D and 3D part-based models. For 2D part-based models, the representation of parts varies from planar patches [19] and 2D ribbons[2], [22] to deformable models[23]. The disadvantage is that it is hard for 2D models to deal with shape variations due to different viewpoints. On the contrary, when 3D data are available, it is possible to match the model directly against 3D data [20]. As these method requires searching through a high dimensional pose parameter space for 3D pose recovery so for a search and rescue mission as the data flow is on-line the calculation time is a critical factor this methods are too time consuming.

- *Cylinder and Super quadric-based methods*
  Bowden et al.[24] encapsulated the correlation between 2D image data and 3D skeleton pose in a hybrid 2D-3D model trained on real life examples. The model they used allows 3D inference from 2D data. The common drawback with the above models is that they do not model the statistical variation among individuals and the effects of clothes on human shape. As the objective of human body detection is to is to individuate the human body in any shape, form, color and size this method can not cover all of the needs for rescue missions.
- *Hierarchical based methods*
  Marr and Nishihara [25] proposed a hierarchical 3D human model. At the highest level of the hierarchy, the body is modeled as a large extended cylinder, which is then resolved into small cylinders forming limbs and torso, and so on to fingers and toes. This method is not sufficient because it contains few actual constraints to support human detection.
- *Contour Based methods*
  Contour-based representations have been used to model the 2D human shape. Baumberg et al. and Sullivan et al[26], [27]. employed a deformable template to handle shape deformation, where the shape model is derived from a set of training shapes. The orthogonal shape parameters are estimated using Principle Component Analysis (PCA). One drawback with this approach is that the model and the extracted contour should be aligned first, which is not a trivial task. Another drawback is that some invalid shapes are produced by the combination of two or more linear deformations. Gavrila et al.[28] Developed a template hierarchy to capture the variety of human shapes, and the model contains no invalid shapes. The common drawback with the above approaches is that they do not model individual parts, and so they can only handle limited shape variety due to articulation

and cannot deal with occlusion very well.

- *Skeleton-based methods*
  Skeleton-based representations[29] have been used to model the topological structure of the human body, but they do not model the shapes of body parts. These approaches are sensitive to noise as a very active factor in rescue scenes and cannot distinguish two classes with the same topological structure but different geometrical structures.

In our experiments in general we use the Contour-based and skeleton-based methods to gathering the information for the next step. while the other methods such as Hierarchical-based methods and Cylinder super quadric-based methods are too sensitive to shape and color variation the part based methodsthe Contour-based applications and skeleton-based methods are flixable enough to shape variation so we consider them the more sufficient for the human body detection.

C. Human body Modeling and Recognition

The third step is the joining and reconstruction of the detected part to find the current position of the body, in the corrent HBD methods there is a straight interaction between the methods of classification and modeling[1]

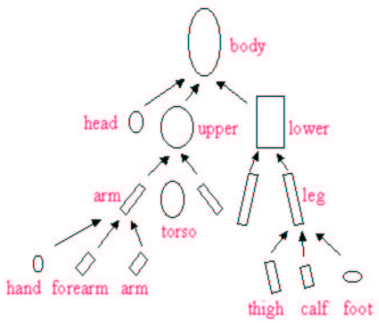

Fig. 1. A human body model

There is a body parts model in Figure 1 that demonstrate the body parts links in some body parts models [1]. There are several methods to detect the composition of the body. The problem of determining the similarity of two shapes has been well studied in several fields. The design of a similarity measure depends on how a shape is represented. Some categories of these methods are:

- *Global shape description*
  Global shape descriptors such as Fourier Transform, moments, and shapes have been used to compare two shapes, but they cannot handle occlusion and local deformation such as articulations very well. The self occlusion o external occlusion are the typical characteristics of the rescue victims this methods are not sufficient for search and rescue missions.
- *Point-based similarity*
  A point-based similarity measure, such as the Hausdorff distance, is commonly used to compare two shapes, but it is very sensitive to noise and occlusion. The common drawback with the above measures is that they have to transform one shape to another before shape comparison because the distance metric is not invariant under similarity transform.
  Various cost functions have been proposed to evaluate the dissimilarity between two contours without aligning them. A cost function weights the similarity of the matched points on the basis of their local properties, such as the difference in the tangent or curvature of the contours at those points. The cost function itself is used to guide the search for the best match. The main drawback of these methods is their high computational complexity due to searching for correspondences at the point level. Furthermore, none of these cost functions is invariant under scaling and/or rotation of the point data.
  Other features such as key points and lines have been used to reduce the computational cost because a digital contour usually consists of much fewer features than of points. Objects are considered similar if their graphs are isomorphic; a similarity metric based on a probability density estimator is used to identify if a shape is an instance of a modeled object.
  To handle occlusion, partial matching is allowed and the largest mutually compatible matches are found by constructing an association graph to search for the maximal clique The main drawback of these methods is that they cannot handle articulated motion because the spatial relationships between features are assumed to be fixed. Moreover, it is difficult to find a coherent set of features that is shared by all possible shapes in a class and that can be extracted reliably.
- *Part-based representation*
  Part-based representations have been proven to handle articulation and occlusion effectively. They have several advantages over other representations such as points, lines, and arcs. First, articulation usually happens at part boundaries, thus, a part-based representation is a more natural and coherent description of articulated shapes. Second, a shape contains fewer aggregate parts than other features. Third, part-based methods find strong support from human vision [1]. The main concerns of a part-based similarity measure are how to decompose a shape into stable parts and how to set up correspondence among them. Parts generally are defined to be convex or nearly convex shapes separated from the rest of the

object at concavity extrema [5], or at inflections [8]. One type of approach is to represent shapes as skeletons or graphs and then to use graph matching or qualitative properties such as topology to compare shapes. The main drawback of these approaches to part-based shape analysis is that the shape decomposition is not stable. Since only qualitative properties are used for shape classification, they cannot distinguish two shapes with the same body part structure but different body part shapes and geometric relationships. Zhu and Yuille [1] developed a similarity measure to compare silhouettes based on both the local shapes of parts and the topology but the method cannot handle shape degeneration or resolution changes very well. Several curve evolution approaches [8] have been proposed to model shapes of an object at different scales, but the related similarity measure is sensitive to occlusion and is not invariant under scaling.

Leung et al. [8] have proposed a method, which combines the intensity pattern and the spatial relationships between the facial features to detect faces from the cluttered environment. However, they do not use the spatial relationship to help detect face features, and no size relationship and recursive procedure is involved in face detection. The main reason is that the facial features have very distinctive patterns and can be detected based on their intensity patterns. In human detection, we rely heavily on the spatial and size relationships to identify the human body parts, because the body parts such as arms and legs do not present very distinctive texture patterns.

IV. Discussion

In order to devise a setting for HBD in the RoboCup Rescue scenario, we have performed a set of preliminary experiments of the techniques proposed in the literature. The goal of these experiments is not an evaluation or a comparison of these techniques, but rather an analysis of their performance when applied in the Rescue scenario.

We present in this section some experiments in the segmentation phase of the process, whose results are depicted in Table I.

In the class of Edge-based Segmentation methods, we have used the Canny method, which is one of the most common method in this class. The computation time for this class of methods is generally good, but the problem of edge grouping is very relevant with these methods, since is related to finding a correct threshold and connecting the one pixel wide-counters. Figure 2 shows also that the edges extracted by these methods are not continuous, and this arises to over segmetation, that may be difficult to dealt with in the next phases of the HBD process.

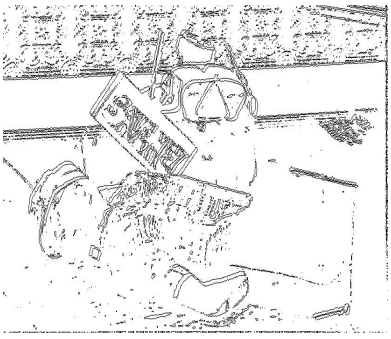

Fig. 2. Edge-based segmentation

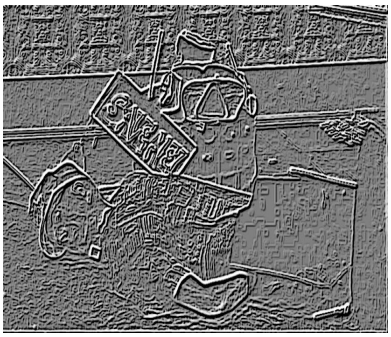

Fig. 3. Clustering-based segmentation

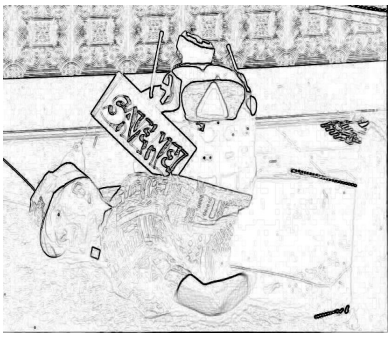

Fig. 4. Region-based segmentation

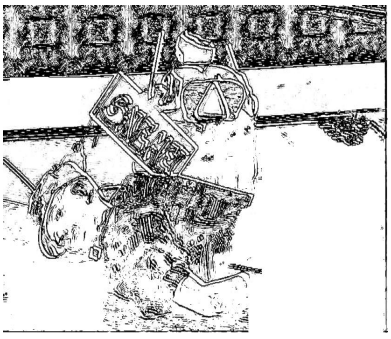

Fig. 5. Split/Merge-based segmentation

Method	Timing	Contour	Threshold	Over-Segmentation
Edge-based	Real Time	Partially Continous	Yes	Yes
Clustering-based	Real Time	Always Continous	No	Yes (Resolve able by merging)
Region-based	Non Real Time	Commonly Continuous	Yes	Yes
Split/Merge-based	Slow Convergence	Commonly Continous	No	No

TABLE I
SEGMENTATION TABLE

Among the Clustering-based approaches, we have used both c-mean and k-mean methods. In this case (see Figure 3) the contours are nearly always closed, however over-segmentation can still occur because the pixels in a cluster are not always adjacent. This problem can be solved by applying a merge process in the homogeneous spaces inside a cluster.

With regard of Region-based approaches, we have experimented both watersheds and pyramidal segmentation methods (see Figure 4). For these methods the computation time is acceptable, while they may cause over segmentation and also the results are sensitive to the use of correct thresholds.

Finally, for the Split/merge-based segmentation, we have used the neighboring regions similarity method. This method does not use any threshold, but the convergence time is very high and in practice it cannot be used on-line. Moreover, as shown in Figure 5, also with this class of methods the contours are not always continuous.

The above results give an indication that there is not a class of methods that clearly overcome the others. In fact, the problem of HBD in the Rescue domain is certainly more complex with respect to other application domains, in which the techniques surveyed in this article have been experimented.

Moreover, the segmentation phase should be designed also according to the next phases of the process, and a general design methodology for developing HBD systems in the Rescue scenario certainly deserves further investigation.

Our future work in this direction includes the analysis and a set of experiments for other methods in HBD applied to the Rescue environment, aiming at devising a HBD system that can be effectively used by an autonomous rescue robot, which is the main objective of our research.

REFERENCES

[1] liang Zhao, "Dressed human modeling, detection, and parts localization." Ph.D. dissertation, The Robotic Institute Carnegie Mellon University,Pittsburgh, 2001.

[2] M. F. D.A. Forsyth, "Body plans," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[3] http://www.informatik.uni-trier.de/ ley/db/conf/vr/vr1993.html.

[4] S. E. U. h, *Computer vision and Image processing: A practical Approach Using CVIPtools*.

[5] L. D. I. Haritaoglu, D. Harwood, "W 4 -real time detection and tracking of people and their parts," Tech. Rep., Aug.1997.

[6] R. rescue, *Robcup Rescue official Manual*, www.robocup.org.

[7] S. Moradi, "Victim detection with an infra red camera ina a rescue robot," 2002.

[8] L. SH.Bahadori, "Human body detection by sensorial information technical report," 2003.

[9] L. V. L. Schomaker, E. Leau, "Using pen-based outlines for object-based annotation and image-based queries," pp. 585–592, 1999.

[10] k. z. l. Shinn-ying ho, "An efficient evolutionary image segmentation algorithm," *Shinn-ying ho,kual zheng lee department of information engineering feng chia university Taiwan*, 2001.

[11] J.F.Canny, "A computational approach to edge detection," *IEEE trans on pattern Analysis and Machine intelligance,Vol 8*, pp. 901,914.

[12] T.N.Pappas, "An adaptive clustering alghorithm for image segmentation," *IEEE trans on Signal processing ,Vol.40 no.4*, pp. 679–698, 1992.

[13] K. J. P. C.W.Chen, J Luo, "Image semgentation via adaptive k-means clustering and knowledge-based morphological operations with biomedical applications," *IEEE trans. Image Processing,vol.7 ,no 12*, pp. 1673–1683, 1998.

[14] Y.L.Chang and X.li, "Adaptive image region-growing," *IEEE trans. on Image processing vol.3 no.6*, pp. 868,872, 1994.

[15] S.A.Hojjatoleslami and J.Kittler, "Region growing:a new approach," *IEEE trans. on Image Processing*, 1998.

[16] D.N.Chun and H.S.Yang, "Robust image segmentation using genetic algorithms with a fuzzy measure," *Pattern Recognition, Vol.29 ,No 7*, 1996.

[17] M.R.Rezaee and P. der Zwet et al., "A multiresolution image segmentation thechnique based on pyramidal segmentation and fuzzy clustring," *IEEE trans. on Image Processing ,Vol.9,No.7*, pp. 1238,1248, 2000.

[18] A. W. P.K.Sahoo, S.Soltani, "A survay of thersholding technique," *CVGIP 41*, pp. 233–260, 1988.

[19] Y. Y. S. Ju, M. Black, "Cardboard people: A parameterized model of articulated image motion," *Proc. Int. Conf. on Automatic Face andvGesture Recognition*, pp. 38–44, 1996.

[20] D. M. Gavrila and L. S. Davis, "3-d model based tracking of humans in action: A multi-view approach," *CVPR'96*, pp. 73–80, 1996.

[21] D. M. I. A. Kakadiaris, "Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection," *CVPR'96*, pp. 81–87, 1996.

[22] E. H. A. S. A. Niyogi, "Analyzing gait with spatiotemporal surfaces," *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 64–69, 1994.

[23] D. M. I.A. Kakadiaris and R. Bajcsy, "Active motion-based segmentation of human body outlines," *Workshop on Articulated Motion*, 1994.

[24] M. S. R. Bowden, T. A. Mitchell, "Reconstructing 3d pose and motion from a single camera view," *BMVC'98*, 1998.

[25] D. Marr and H. Nishihar, "Representation and recognition of the spatial organization of three-dimensional shape," *Proc. Roy. Soc. B, B-20*, 1977.

[26] D. H. A. Baumberg, "Learning flixable models for image sequences," *Proc. Europian Conf. on Computer Vision*, 1994.

[27] e. M.Sullivan, C.Richards, "Pedesterian tracking from stationary camera using active deformable models," *Proc.Intelligent Vehicles*, 1995.

[28] D.M.Gavrila and Philomin, "Real-time object detection for smart vehicles," *Int'l Conf.On Computer Vision*, 1999.

[29] S. andA.L.Yuille, "Forms: A flexble object recognition and modeling system," 1996.