Visual Fall Detection for Intelligent Spaces

José M. Cañas, Sara Marugán
Robotics Group
Univ. Rey Juan Carlos
28933 Móstoles
jmplaza@gsyc.es, s.marugan@alumnos.urjc.es
http://www.robotica-urjc.es

Marta Marrón, Juan C. García
Dept. Electrónica
Univ. Alcalá
28871, Alcalá de Henares
marta,jcarlos@depeca.uah.es
http://www.depeca.uah.es

***Abstract*— Artificial vision provides a remarkable good sensor when developing applications for intelligent spaces. Cameras are passive sensors that supply a great amount of information and are quite cheap. This paper presents an application for elderly care that detects falls or faints and automatically triggers the health alarm. It promotes the independent lifestyle of elder people at their homes as the monitoring application will call for timely health assistance when needed. The system extracts 3D information from several cameras and performs 3D tracking of the people in the intelligent space. It uses an evolutive multimodal algorithm to continuously estimate the 3D position in real time and learns the visual appearance of the persons. The system has been validated with some experiments in different real environments.**

I. Introduction

Over one-third of elders 65-years-old fall each year [8]. The falls usually result in serious injuries like hip fracture, head traumas, etc. The rapid health assistance in case of fall may reduce the severity of the injuries. The care of elderly implies a continuous monitoring of their daily tasks. In many cases their own families or the social services are in charge of their care at their own homes or in specialized institutions. But even counting with the necessary amount of caregivers, it is impossible to watch these patients continuously in order to detect any incident as fast as possible. The problem worsens for people who live alone at home, as they need much more this type of assistance in case of emergency.

In the context of fall detection and prevention there are several technological products in the market. First, traditional monitoring systems as pendants or wristbands worn by the patients [6], who must activate such devices when needed, usually pressing a button. The system sends an emergency call to the appropriate health service. These traditional systems require human intervention to report an alarm or ask for help, and user's potential non-compliance (both intended and unintended) is a potential problem. In certain situations, for instance a faint that causes a fall to the floor, it will not be possible for the patient to activate the device, and that can be dangerous as the severity of the damage may increase with the time at the floor without health assistance. A second group of wearable systems relies on accelerometers and tilt sensors to automatically detect the falls [9]. Carrying this devices continuously may become a nuisance for the users.

Other solutions are embedded in the environment, they use external monitoring devices and then, the user's compliance is not required. There are systems which are based on floor-vibrations [12], on infrared array detectors [11] and on cameras. Several vision based systems use omni-directional cameras [10], [7]. In particular [7] looks for activity patterns, models the patient's behavior and detecting abnormal activities. Other works use optic flow as the main visual feature [14] or the motion history and human shape variation [13].

In this article we introduce a system with a set of regular cameras that monitors the patient movement automatically. When it watches some anomalous patient behavior such as a falling to the floor, the system automatically can send an emergency call for immediate health assistance. The system extracts the three-dimensional position of people at the scene and tracks them from the video streams of several cameras.

II. Fall detection application based on 3D position

For monitoring applications, a great part of the useful information in the work space is mainly three-dimensional, like the relative position of an object opposed to another or the movement of a person. One of the main problems in the identification of dangerous situations using vision sensors is their two dimension nature. For instance, when using a flat image to detect whether a person is near to an ignited oven, a window, a door, etc or not, there is ambiguity in the estimation of the distance, so we could easily make a mistake. High risk situtions are better described, and in a more simple way, in 3D spatial terms.

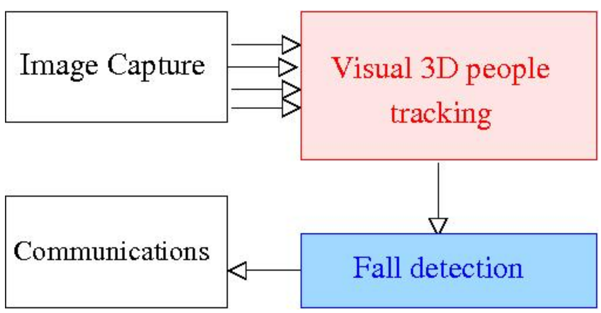

Fig. 1. Block diagram of the developed application

We have built an application, named *eldercare*, whose main blocks are depicted at Fig. 1. First, the *image capture* block is responsible of getting the frames from several camera sensors along the monitored area. Analog cameras, wireless cameras, firewire and regular USB cameras are supported. Second, the *visual 3D people tracking* block extracts three-dimensional information in real time from the images, tracking the 3D position of every person at the monitored zone. It provides the current 3D position of every person at the area to the *fall detection* block. This third block defines a set of alarm rules which take into account 3D position and time conditions to trigger a health assistance alarm. For instance, if the position of a person is close to the floor (less than 20 cm) for a minute or more then the fall condition is triggered and an alarm is signaled to the *communications* block. This fourth block is responsible of sending such alarm to the health services via SMS, MMS, automatic phone call, etc.

In addition, a graphical interface has been developed, but only for debugging purposes. The system itself presents no window at operation time and records no single image to keep privacy of the monitored people. It has been implemented with a set of low cost cameras and a conventional PC. The 3D estimation technique has been carefully designed to run at real time of commodity hardware and it will be described in detail in section III.

III. Evolutive vision algorithm for 3D people tracking

An evolutive multimodal algorithm has been developed to track the 3D people position in real time. The system needs several (two, three of four) calibrated cameras along the monitored area and uses simple visual features like color and motion to keep track of the persons.

The evolutive algorithm manages a population of hundreds of individuals, which are possible person locations (X, Y, Z). Each individual is a 3D hypothesis and has a fitness associated which indicates its likelihood of being a good hypothesis, that is, a real 3D person location. Such fitness is computed from the images using color and motion in the frames, as shown in section III-C.

The algorithm starts with a random population of individuals scattered along the 3D monitored volume. Then repeatedly runs two steps until the populations converge to the problem solutions. The first step is the fitness computation. Individuals with high fitness value have more chances of passing to the next generation. The second step is the generation of a new population using genetic operators as random mutation, elitism, thermal noise, crossover and abduction from current images. Hopefully the evolution and the improvement of the population achieve new generations that are increasingly closer to the problem solution, that is, that better capture the 3D people locations.

The algorithm resembles the particle filters applied in other vision-based tracking systems [3], [4], but does not hold the Bayes and MonteCarlo probabilistic requirements and so it does not have the convergence assured. Instead of particles we have individuals, instead of posterior probability the individuals have fitness. The algorithm can be seen as a search algorithm in 3D euclidean space, and has the flexibility of the evolutive algorithms where ad-hoc operators and fitness functions can be developed to improve the performance of the system. In constrast with many works in the visual tracking literature, the proposed system does not track the persons in the images neither merges the 2D estimations into a 3D one. It looks for people directly in 3D, and the images are used only as observations which feed and validate the 3D hypothesis. This way the 2D clustering is avoided.

A. Exploration race

The whole population of the evolutive algorithm is divided into an *exploration race* and several *tracking races*. The exploration race looks for persons in the monitored area performing the coarse-grained search, and each tracking race follows one single person, performing the exploitation or fine-grained search of the algorithm.

The *explorer race* consists of 400 3D individuals which search for movement in the whole space covered by cameras. At each iteration of the algorithm a new explorer population is generated through random mutation and abduction genetic operators. Random mutation creates hypothesis sampling a uniform probability distribution between the X_{min} and X_{max}, Y_{min} and Y_{max}, Z_{min} and Z_{max} of the work space.

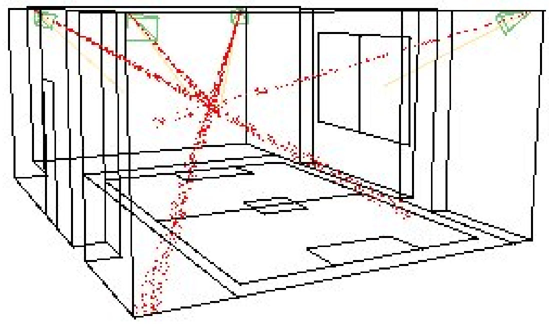

Fig. 2. Abduction operator generates individuals in the 3D backprojection ray of pixels where motion has been detected

As can be seen in Figure 2, abduction operator puts new individuals in 3D areas in the projection line of pixels where image motions has been detected. First, images are filtered looking for pixels with motion, as it will be explained at section III-C, mainly using background substraction. These motion pixels are sampled, and for each selected one the 3D ray which projects into such pixel is computed using known extrinsic and intrinsic parameters of the corresponding camera. Several new explorer individuals are located in 3D along that ray, sampling a uniform distribution between 0 and the maximum depth of the ray (it depends on the work space). This way the explorer search is not so blind, and its convergence to interesting 3D areas is speeded up.

In the case of detecting a 3D area with significant motion, a new tracking race will be created around that location if no current tracking race covers it. The explorer individuals are evaluated using the motion fitness defined at section III-C and only those above a fitness threshold are allowed to generate a new tracking race. In addition, those explorers close to the 3D position of an existing tracking race are inserted into such race if their fitness is better than the worst current individual of such race. This way new races are created only in areas found by the explorers that are not covered by any current tracking race.

B. Tracking races

Each race is dedicated to the 3D tracking of a single person and it is composed of around 50 3D individuals. The algorithm achieves multimodality using one tracking race per person, so in the case of four people at the room, the algorithm will create four races, as in Figure 3. Each one learns the visual appearance of the person it is following, as will be explained at III-D. It also accounts for the spatial movements of such person and its 3D hypothesis follow such displacements. All the individuals of a race are continuously validated in all the cameras computing their fitness. They are projected into the camera images and the algorithm checks whether such projecting pixel lies in a patch with the race color or not and checks whether it lies in a patch where motion has been detected. The individuals are ordered by fitness and the race 3D position is computed as the weighted sum using the fitness as weight. This way only the individuals which compatible with the images define the 3D person position estimation, without bias of the bad individuals.

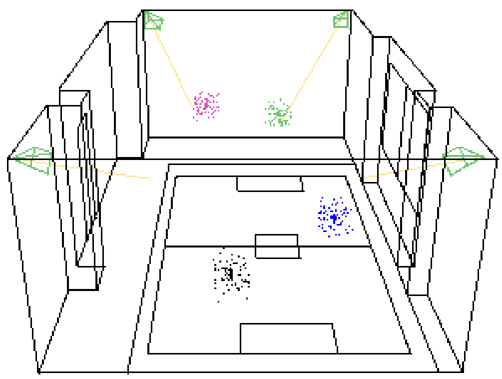

Fig. 3. One tracking race is generated per person in the monitored area

New generations of the tracking races are generated by elitism and thermal noise operators. By elitism, the best individuals directly pass without change to the next generation as they are good estimations of person position so far. Thermal noise operator adds gaussian 3D noise in X-Y-Z coordinates to the position of the individuals and generates new ones in the vicinity. This operator is useful to follow the person movement, as it will generate individuals in the new position of the person as she moves in the monitored area. It will also generate new hypothesis in bad positions around the current one, but these will be descarded as they will not get enough fitness to survive. The particular number of individuals generated by elitism and by thermal noise on each iteration is a parameter of the system, and must be tuned.

The races are born from the most promising explorers, as previously seen, but they have to gather enough fitness to survive. If the best individual of a tracking race does not reach a fitness threshold that means that it has lost its person, and such race disappears. This way, the number of tracking races is dynamic, and so the number of total individuals in each generation. There always be one exploration race (400 individuals) and the number of tracking races (50 individuals each) depends on the number of persons in the monitored area. At the beginning there is no one but as soon as a person shows up in the scene the algorithm will create a new race to track her 3D position.

C. Fitness function from images

Fitness computing is composed of *motion fitness* and *color fitness*. The explorer quality is calculated based only on motion information. In contrast, once a tracker race has been initiated, the system learns automatically the color of its person and its clothes. Then, a color fitness is also computed for each of the individuals of the tracking race, and both visual clues, motion and color, are taken into account.

1) Motion fitness and motion detection: For each explorer individual i its fitness h_i is defined in equation (1), where the summatory extends to all the cameras of the system and $P(mov_i|img_m)$ is computed as follows. The 3D individual $i = (X, Y, Z)$ is projected on camera m and so, Pix_i is computed. The number of pixels k with motion in a 5x5 neighborhood around Pix_i is calculated and then the equation (2) applied.

$$\mathrm{h}_i = \sum P(\mathrm{mov}_i \mid \mathrm{img}_m) \tag{1}$$

$$P(mov_i|img_m) = \frac{max(1,k)}{25} \tag{2}$$

A motion image is computed at every iteration of the algorithm to compute this motion fitness. When an object is in movement, some pixels change their values. The system detects these changes through comparison of consecutive frames and through comparison with the learned background. The background is estimated for each camera of the system as a weighted sum of frames at defined intervals (see equation 3), where α is in range [0,1] (0.2 in our settings) and β indicates time interval for background updating (4 minutes in our settings). Frame comparison is just the absolute difference between two consecutive images. If a pixel difference is above a predefined threshold, that pixel passes the motion filter as it has a significant difference with regard to previous frame or background image. Figure 4 displays an example of

Fig. 4. Input image from one of the cameras (left) and computed motion image (right).

motion image computed comparing the current frame with the previous one and the learned background image.

$$backgrd(t) = \alpha\, backgrd(t-\beta) + (1-\alpha) frame(t) \quad (3)$$

2) Color fitness: Now let be i the individual of a tracking race, its fitness h_i is calculated following the equation (4). $P(\mathrm{mov}_i \mid \mathrm{img}_m)$ is computed in the same way as for explorers and $P(\mathrm{color}_i \mid \mathrm{img}_m)$ is computed as follows. The i 3D individual is projected on camera m and so, Pix_i is computed. The number of pixels c inside a 5x5 neighborhood around Pix_i that matches the color filter for that tracking race is calculated and then equation (5) applied.

$$\mathrm{h}_i = \frac{1}{2}\left(\sum P(\mathrm{mov}_i \mid \mathrm{img}_m) + \sum P(\mathrm{color}_i \mid \mathrm{img}_m)\right) \quad (4)$$

$$P(color_i|img_m) = \frac{max(1,c)}{25} \quad (5)$$

Fig. 5. Input image (left) and image filtered with the color of one tracking race (right).

Figure 5 shows an input image filtered with the color of one of the tracking races. This filtered image will serve to compute the color fitness of all the individuals of that race. Each race has its own color description in the HSV color space, as it will be detailed in section III-D.

D. Learning of color

People tracking based only on movement detection does not solve the problem when the person stops and remains still. *Eldercare* system learns object color to keep the track in this situation, because people usually sit down for a long time in their daily lives and they must not disappear in the monitoring application.

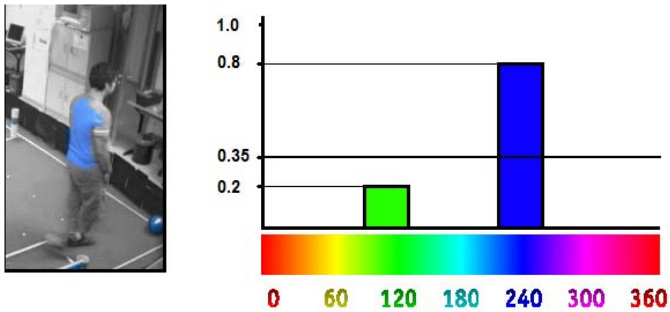

Fig. 6. Color learning based on H histogram

The representative color of a race is stored in Hue-Saturation-Value (HSV) color space, as H_{race}, S_{race} and V_{race} plus one tolerance on each color channel. The system uses motion filter to learn the person color at the very beginning. When the new race is created, the system takes all the pixels passing the motion filter as samples for building a HSV histogram, which is associated to the tracking race. The bins of the HSV histogram with enough samples are taken as the definition of the race color. For instance, Figure 6 shows the part of an input image where most of the explorers project, and how their color mainly fall in the blue bin. That is the only bin with a percentage above the threshold (35%), and so, the only one that defines the color of that race, which will cover the blue T-shirt of the person.

To cope with changing lightning conditions the color of a race is periodically redefined from the 3D individuals of the race with high fitness, using again the method described above.

IV. Experiments

Several experiments have been carried out to validate the proposed system in real scenarios. The first set of experiments was run in a Pentium IV at 2.6 GHz, with 512 MB of RAM memory. The Robotics Lab at the Universidad Rey Juan Carlos (7m x 4m x 3m) was equipped with four iSight cameras close to the ceiling, one at each corner, as shown in Figure 2. They provide 320x240 pixel images at 15 fps. One of them was connected to the main computer directly through the local firewire bus and the other three through ethernet, using three video servers at the same room.

Figure 7 shows one of the input images (left), the color filters learned (center) and the 3D individuals of the two tracking races generated to track the two persons in the room (right). There is no problem in both persons wearing similar clothes as long as they are in separated positions, their color descriptions are learned independently. If both persons walk together, side by side, the system will treat them as a single person but as soon as they split again a new race is created and they are tracked as two different persons.

Fig. 7. Two races and their color description learned.

In the experiment shown at Figure 8 the system tracked three people simultaneously and their 3D trajectories are displayed. People entered into the lab at different times and the system incorporated them as soon as they showed up. Trajectories and 3D individuals are displayed in the learned color for the corresponding person. The system tracked them without difficulty keeping 12 iterations per second for the evolutive algorithm, fast enough to follow the people 3D movements in real time. The 3D skeleton of the environment must be given to the application, which uses it for visualization purposes in the debugging graphical interface, to let the human easily understand the generated 3D trajectory data.

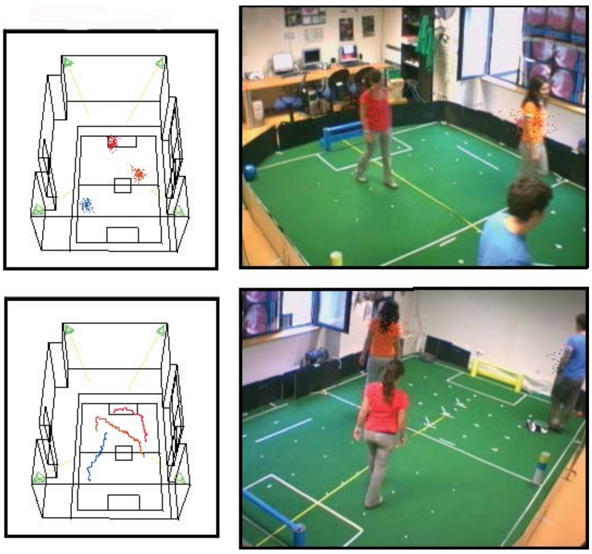

Fig. 8. Typical execution, 3D tracking of several people.

Finally, the system was succesfully tested on a technological fair at the Centro Nacional de Tecnologías de Accesibilidad (León, Spain). The space for the demo was equipped with four Axis network cameras connected through ethernet to a MiniMac computer. In Figure 9, the system has detected that one person wearing a red T-shirt has fallen to the floor and the alarm has been triggered (red label on the left). His trajectory in 3D is also displayed (in the learned color, red) in the left part of the figure, together with the 3D skeleton of the room. The person fall can also be observed in that trajectory.

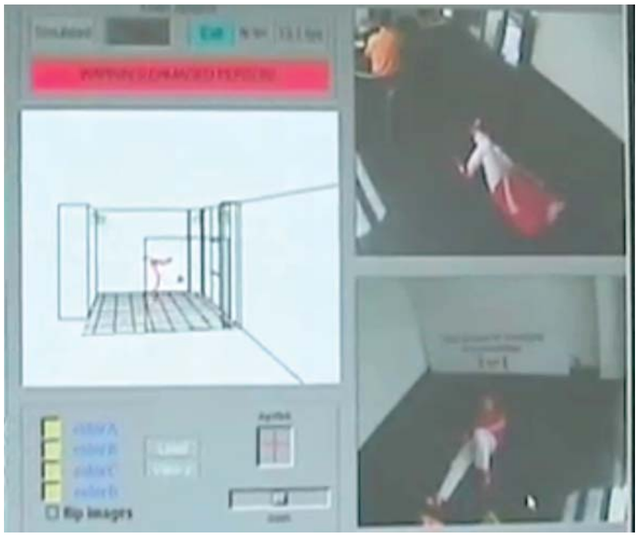

Fig. 9. Person lying on the floor, the fall has been detected.

V. Conclusion

We have presented a system, named *eldercare*, that automatically detects falls of elder people in an intelligent space and triggers emergency alarm. It may be useful for elderly living alone at their homes, who may enjoy both the independent lifestyle and the safety of immediate health assistance in case of a fall or faint. In contrast with other traditional fall detection systems, like pendants of wristbands, *eldercare* is passive and doen't require any user action or wearing specific devices. It consists of several regular cameras and a commodity personal computer that runs the image processing software. No single image is stored to preserve the user privacy.

The proposed evolutive multitarget tracking algorithm uses color and motion as the main visual clues in its fitness function, and continuously estimates the 3D position of people in the monitored area. The fall situation is defined in terms of 3D positions close to the floor and time conditions. The individuals of the evolutive algorithm are 3D points organized in an exploration race and several tracking races, one per person.

The proposed algorithm has been validated building a prototype and performing real experiments in two different real scenarios. The system is able to properly track the 3D position of several people in real time, as seen in section IV. In the experiments, the error in the 3D position estimation has always been below 15 centimeters, good enough for this fall detection application.

We are working to introduce new visual features like SIFT, SURF etc. into the tracking algorithm. They hopefully will be more robust than color descriptions of the tracked persons. We also intend to expand the system with more cameras to cover more than one room, and to detect other dangerous situations like proximity of Alzheimer patients to risky areas (the exit door, the windows, etc).

References

[1] Barrera P., Cañas J. and Matellán V., "Visual object tracking in 3D with color based particle filter", *Int. Journal of Information Technology*, Vol 2, Num 1, pp 6165, 2005.

[2] Fritsch J., Kleinehagenbrock M., Lang S., Fink G. and Sagerer G., "Audiovisual person tracking with a mobile robot", *Proceedings of Int. Conf. on Intelligent Autonomous Systems*, pp 898906, 2004.

[3] Pérez P., Vermaak J. and Blake A., "Data fusion for visual tracking with particles", *Proceedings of IEEE*, Vol 92, Num 3, pp 495513, 2004.

[4] Pupilli M. and Calway A., "Real-Time Camera tracking using a particle filter", *Proceedings of British Machine Vision Conference*, pp 519528, 2005.

[5] Zotkin D., Duraiswami R. and Davis L., "Multimodal 3D tracking and event detection via the particle filter", *IEEE Workshop on detection and recognition of events in video*, pp 20-27, 2001.

[6] Brownsell S., Hawley M.S., "Fall monitoring", *Home telehealth: connecting care within the community*, Eds: R.Wootton, S. Dimmick y J. Kvedar, pp 108-120, Royal Society of Medicine Press, 2006.

[7] Miaou S.G, Shih F.C., Huang C.Y., "A Smart Vision-based Human Fall Detection System for Telehealth Applications", *Proc.Third ISATED Int. Conf. on Telehealth*, pp: 7-12, Montreal, Canada, Ed: R. Bashshur, Acta Press, 2007.

[8] Rajendran P., Corcoran A., Kinosian B., Alwan M., "Falls, Fall Prevention, and Fall Detection Technologies", *Eldercare Technology for Clinical Practitioners*, pp: 187-202, 2008 Humana Press.

[9] Yang C.C., Hsu Y.L., "Developing a Wearable System for Real-Time Physical Activity Monitoring in a Home Environment", *Proc. Third ISATED Int. Conf. on Telehealth*, Montreal, Canada, Ed: R. Bashshur, Acta Press, 2007.

[10] Yiping T., Shunjing T., Zhongyuan, Y., Sisi Y., "Detection Elder Abnormal Activities by using omni-directional vision sensor: activity data collection and modeling", *Proc. Int. Joint Conference SICE-ICASE*, pp: 3850 3853, 2006.

[11] Sixmith A., Johnson N., "A smart sensor to detect the falls of the elderly", *Pervasive Computing*, vol. 3 no. 2, pp: 42-47, 2004.

[12] Majd Alwan, Prabhu Jude Rajendran, Steve Kell, David Mack, Siddharth Dalal, Matt Wolfe, Robin Felder., "A Smart and Passive Floor-Vibration Based Fall Detector for Elderly", *The 2nd IEEE International Conference on Information & Communication Technologies: from Theory to Applications - ICTTA06*, April 24 - 28, 2006 in Damascus, Syria.

[13] Rougier C., Meunier J., St-Arnaud A., Rousseau J., "Fall Detection from Human Shape and Motion History Using Video Surveillance", *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, vol. 2, pp.875-880, 2007.

[14] Huang C.L., Chen E.L , Chung P.C., "Fall Detection using Modular Neural Networks and Back-projected Optical Flow", *Biomedical Engineering - Applications, Basis and Communications*, Volume 19, No. 6, pp. 415-424, 2007.