

# Which Strategy to Combine Face Identification Tools with Clothing Similarity: Contesting or Reinforcing?

Saïd Kharbouche and Michel Plu

Orange Labs 2, Av Pierre Marzin, 22307 Lannion, France

**Abstract.** This paper describes a novel and efficient approach that integrates clothing similarity into face identification process in personal photos. The information extracted from people's clothes would be helpful if they are dissimilar, however, this information could make errors and noise if we have some people with similar clothes. To resolve this problem, we propose here a better methodology that exploits clothing similarity. The main idea is summarized as follows: if a person is well identified in a detected face, instead to reinforce this person in every face (in other photo) with similar clothes, we contest her/him in every face with dissimilar clothes. The weight and the influence of the information extracted from a face in a photo to another face depend on the spatiotemporal distance between photos, the similarity degree between the clothes and the incertitude level about their real identities. We utilize belief functions theory in order to manage efficiently the imprecision and the uncertainty. Besides, the results obtained showed off the interest of our approach.

## 1 Introduction & Previous Work

Every user of digital photo technologies regularly found himself with large collection of images to annotate for browsing and later photo retrieval. So, lot of systems have been conceived with the aim of helping users to index and annotate their personal photos. Some systems are web based, as they permit the users to share online the photos with their communities. Also, other systems are accessible by mobile devices (GSM, PDA,...) which allow users a realtime photo sharing [6, 9, 13]. In order to retrieve these shared photos, there are many ways to organize them but the most current ways are:

- Indexing by dates: time-taken;
- Indexing by location: place-taken;
- Indexing by people's identities: photographed people.

The date and location are encoded in photo as a timestamp of creation date and GPS coordinates, so the indexing by date or location can be an automatic and quit reliable process. Also in the case of cameraphone, the radio cell id can be used as local identification [6]. But the indexing by people's identities is very difficult to make it as an automatic process. In personal photos the faces have various sizes, positions and image qualities. To resolve people's identities in photo, some systems use contextual information (e.g. date, local, events, comments analysis, social data, co-presence detection

using bluetooth...) either lonely or combining with face recognition outputs [6, 12].

Clothes appearance has been investigated in many work. The authors in [1] describe a technique for building a composite photo that represents an event. Their approach classifies the photos in four classes using K-NN algorithm and clothing similarity (color and texture). The clothes are region of upper body that can be localized by a face detection. The representative photo of a folder includes four sub photos of faces (cluster centers) and a main photo that contains the maximum number of faces. In [18] a technique of semi-automated face annotation was detailed. The proposed technique combines the both recognition results (face recognition, and clothing similarity), and deals with missing features of face or clothes. The latter is named contextual features and they are extracted from the extended face region. The authors used a fusion technique based on conditional probability. In [3], a system of face semi-automated annotation combines the face similarity, clothing similarity and uniqueness constraint (no person appears in the same photo twice) by a loopy belief propagation (LBP). The clothing similarity is used for the photos of the same event (interval of 4 or 6 hours) and clothing features (color and texture) are extracted from the upper body region. The position of this region is relative to the detected frontal face. In this last work, in order to avoid the errors engendered by clothing similarity, the latter is less weighted compared to face recognition. The paper [16] describes a personal photos clustering approach based on face recognition and clothing similarity. The clothes pieces are localized by face detector, and the vector feature is based on descriptors of colors and texture. In addition, it proposes a method for skin detection that exploits face color, and the detected skin will not be used for further computing. The scores of face recognition and clothing similarity are combined linearly, so the result score is integrated in logistic function (likelihood measuring) in order to compute the similarity between faces. The uniqueness constraint is integrated in the clustering algorithm (K-means) as a hard constraint.

The aim of our system is to suggest the metadata list (people's identities) of each detected face in the image. Given two faces which are detected in two different photos with a small spatiotemporal gap, then their clothes features are extracted. In previous work [18, 3, 16], the use of clothing similarity exploits essentially this idea: If clothing similarity is high, the identities of faces are probably the same. So, for a face to identify, every person who has already been detected with similar clothes in other photo of the same event (same date and place) will be reinforced, thus pushed forward in the candidates list. This method is called reinforcing strategy. Another strategy (our contribution in this paper) is called contesting strategy that exploits the following idea: if clothing dissimilarity is high, the identities of faces are probably different. So, for a face to identify, every person who has already been detected with dissimilar clothes in other photo will be contested, thus pushed back into candidates list. The contesting strategy is more logic than reinforcing strategy because, photographed people would be with similar color and texture of their upper bodies (similar clothes and skin). Then, we can't identify with certainty a face only via its clothing similarity. However, the use of clothing similarity could provide us with certainty the non-matching between a detected face and a given person identity that we know his/her clothes. In [16], the clothing similarity becomes misleading in the case of people with similar clothes. Also, for this reason, the

clothing similarity is less weighted than face recognition in [3]. The reinforcing strategy results important errors in the case of people with similar clothes.

In our previous work [11], we proposed a method based on belief functions for combining face detection, face recognition and gender recognition, and we introduced a logic-based information named Inter-Zones (or uniqueness constraint). The later consists to contest every person who has already recognized in another face of the same photo. The thesis [10] details more our system and it explains how we have combined the information extracted from audio-textual comments, spatiotemporal data and the information extracted from image analysis tools. Our personal-photos indexing system was integrated in a web-based system "Someone" [2] of multimedia documents indexing and sharing, and it includes two major information sources:

**IA: Image Analysis** it includes three tools: Face Detection  $-FD-$ , Face Recognition  $-FR-$  and Gender recognition  $-GR-$ ;

**CA: Comments Analysis** it consists of recognizing the people who were quoted in textual and vocal comments.

However, in our precedent work, we identify people's identities in photo without taking into account the information that could be extracted from clothes.

In this paper we won't use "comment analysis"  $CA$ , but we will focus on "images analysis"  $IA$  and a tool of Clothing Similarity  $CS$ . So, we integrate the information issued from others photos of the same collection for face identification process. Thereby, each face in a given photo will take part in the decision for resolving the identity of all faces of other photos. It depends on spatiotemporal distance between photos, image analysis ( $IA$ ) results and the clothing similarity. Belief functions theory has been chosen in our fusion process in order to manage imprecision and uncertainty. Many works attempt to compare the formal framework of this theory to several other methods of information fusion, like the probabilistic methods, the possibility theory and to linear combination. For the multimodality fusion, the belief functions are introduced in many systems, and it was compared with success to many others techniques of fusion like possibility and linear combination [4].

This paper is organized into the following sections. Second section gives a summary of belief function theory, particularly the rules that will be called in our proposed approach. Third section describes our fusion process, followed by a section that gives our evaluation and preliminary results. Lastly, we draw our conclusion.

## 2 Belief Functions Theory

Here, we recall basics of this theory, especially the tools that we will use in our fusion process. *Dempster's* work on lower and upper limits of probability distributions [7] allows *Shafer* to build the fundamental of belief functions theory [14].

Let  $X$  be an unknown form and,  $\Psi$  a non empty finite set of  $q$  assumptions (about the reality of  $X$ ), also called the frame of discernment:

$$\Psi = \{h_1, h_2, \dots, h_q\}. \quad (1)$$

Let  $S$  be an information source (or piece of evidence) which gives data about the membership between  $X$  and the elements of  $2^\Psi$ . A basic probability assignment or mass function  $m_S^\Psi[X]$  maps  $2^\Psi$  to the interval  $[0, 1]$  and satisfies this condition:

$$\sum_{A \in 2^\Psi} m_S^\Psi[X](A) = 1. \quad (2)$$

$m_S^\Psi[X](A)$  represents the degree of belief attributed exactly to the subset  $A$  of  $\Psi$ , this can't be divided between the elements of  $A$ . The belief function  $bel_S^\Psi[X]$  maps  $2^\Psi$  to  $[0,1]$ , it can be derived from the mass function using this formula:

$$bel_S^\Psi[X](A) = \sum_{B/B \subseteq A} m_S^\Psi[X](B) \quad \forall A \in 2^\Psi. \quad (3)$$

This measure may be interpreted as the minimum degree of belief attributed to  $A$ .

## 2.1 Data Combining Rules

Given two information sources  $S_1$  and  $S_2$  (or two pieces of evidence). To identify  $X$ , we derived two mass functions  $m_{S_1}^\Psi[X]$  and  $m_{S_2}^\Psi[X]$  from  $S_1$  and  $S_2$  respectively. These functions can be incorporated in one mass function by the conjunctive combination rule, noted by  $\odot$ . It is defined for all  $A \in 2^\Psi$  as follows:

$$m_{S_3}^\Psi[X](A) = (m_{S_1}^\Psi[X] \odot m_{S_2}^\Psi[X])(A); \quad (4)$$

$$= \sum_{B \cap C = A} m_{S_1}^\Psi[X](B) \cdot m_{S_2}^\Psi[X](C). \quad (5)$$

Hence, the degree of conflict between  $S_1$  and  $S_2$  is the mass attributed to  $\emptyset$ . In order to work in a closed world so that  $\Psi$  must be exhaustive (include all possible assumptions) and the mass attributed to  $\emptyset$  must be null, we use *Dempster's* rule of combination noted by  $\oplus$ , also called orthogonal rule. It is defined for all  $A \in 2^\Psi - \emptyset$  as follows:

$$m_{S_4}^\Psi[X](A) = (m_{S_1}^\Psi[X] \oplus m_{S_2}^\Psi[X])(A); \quad (6)$$

$$= \frac{m_{S_3}^\Psi[X](A)}{1 - m_{S_3}^\Psi[X](\emptyset)}; \quad (7)$$

$$m_{S_4}^\Psi[X](\emptyset) = 0. \quad (8)$$

*Dempster's* rule is possible only if  $S_1$  and  $S_2$  are not in total conflict, in other words if  $m_{S_3}^\Psi[X](\emptyset) < 1$ . The function  $m_{S_1}^\Psi[X](\Psi) = 1$  is a neutral element of " $\oplus$ ":

$$m_{S_1}^\Psi[X] \oplus m_S^\Psi[X] = m_S^\Psi[X] \quad \forall S. \quad (9)$$

And the function  $m_{S_1}^\Psi[X](h_i) = 1$  is an absorber element of " $\oplus$ ":

$$m_{S_1}^\Psi[X] \oplus m_S^\Psi[X] = m_{S_1}^\Psi[X] \quad \forall S, \text{ with } m_{S_1}^\Psi[X] \odot m_S^\Psi[X](\emptyset) < 1. \quad (10)$$

## 2.2 Decision Making

There are several kinds of decision rules employed in the evidence theory. One of those methods was proposed by *Ph. Smets* [15]. It transforms the evidence model to a probability model by a function noted  $BetP_{m_S^\Psi[X]}(\cdot)$ . It is called pignistic probability function, which is formalized for all  $h_i$  as follows:

$$BetP_{m_S^\Psi[X]}(h_i) = \frac{1}{1 - m_S^\Psi[X](\emptyset)} \sum_{A \in 2^\Psi / h_i \in A} \frac{m_S^\Psi[X](A)}{|A|}. \quad (11)$$

$|A|$  represents the cardinality of  $A$ .

## 2.3 Processing Time

Evidently, working with  $2^{|\Psi|} = 2^q$  sets of  $\Psi$  expands the fusion processing time. In order to resolve this problem we have developed an algorithm that deals only with non-null values of mass function (focal elements). This algorithm will be explained in a future paper.

## 3 Fusion Process

This section explains how we integrate the clothing similarity in the process of people's identities identification in personal-photos. We will try to combine the result of clothing similarity ( $CS$ ) with one of our current image analysis  $IA$  using two strategies: contesting and reinforcing. To start with, we will extract two mass functions from the information sources  $IA$  and  $CS$ . Then we combine them into a unique mass function for metadata (people's identities) relevant measuring.

### 3.1 Notations

We introduce now some basic annotation that we will use in the rest of this paper. Given a user  $U$  and its address book  $B$ , the later represent a set of  $N$  known people to the user  $U$ . Each person is represented in this set by: its names (last name, first name and nicknames), its face images, its gender and its id.

$$B = \{P_1, P_2, \dots, P_i, \dots, P_N\}. \quad (12)$$

Then, to identify any face we use these exhaustive and exclusive discernment frame:

$$\Omega = \{P_1, P_2, \dots, P_i, \dots, P_N, UM, UF, noFace\}. \quad (13)$$

$UM$  and  $UF$  denote unknown male and unknown female respectively. This frame is exhaustive and exclusive because the reality of each detected face is an unique assumption of  $\Omega$ .  $I$  symbolizes a photos collection:

$$I = \{I_1, I_2, \dots, I_s, \dots, I_S\}. \quad (14)$$

Each photo  $I_s$  is accompanied with two metadata, photo-taken time  $t_s$  and photo-taken place coordinates (GPS coordinates).  $F_s$  denotes the set, if the faces that have been detected by  $FD$  (face detector) in photo  $I_s$ :

$$F_s = \{F_{s,1}, F_{s,2}, \dots, F_{s,f}, \dots, F_{s,z(s)}\}. \quad (15)$$

$z(s)$  is the number of faces that have been detected in photo  $I_s$ . So, we determine the clothes detached to these faces:

$$R_s = \{R_{s,1}, R_{s,2}, \dots, R_{s,f}, \dots, R_{s,z(s)}\}. \quad (16)$$

$R_{s,f}$  denotes a part of the clothes attached to the face  $F_{s,f}$  (a part of upper body region).

### 3.2 Images Analysis: IA

We seek to identify a face  $F_{s,f}$ . So, for this face we have four information sources: face detection ( $FD$ ), face recognizer ( $FR$ ), gender recognizer ( $GR$ ) and uniqueness constraint ( $IZ$ ). All these information are aggregated in  $IA$  as follows:

$$m_{IA}^{\Omega}[F_{s,f}] = m_{FD}^{\Omega}[F_{s,f}] \oplus m_{FR}^{\Omega}[F_{s,f}] \oplus m_{GR}^{\Omega}[F_{s,f}] \oplus m_{IZ}^{\Omega}[F_{s,f}]. \quad (17)$$

For more detail we refer the reader to [11].

### 3.3 Clothing Similarity: CS

In photos collection  $I$ , there are  $r$  detected faces, thus  $r$  associated clothing parts (with  $r = (z(1) + \dots + z(S))$ ). Each face and its clothing in a photo different to  $I_s$  will be considered as an information source for identifying the face  $Z_{s,f}$ , consequently we obtain  $r - z(s)$  information sources.

**Contesting Strategy.** Always, we try to identify the face  $F_{s,f}$ . Let  $F_{s',f'}$  be another face (with  $s \neq s'$  and  $f' \in \{1, \dots, z(s')\}$ ) and  $R_{s',f'}$  be its related clothes, the masses function issued from  $R_{s',f'}$  about the metadata  $P_i$  can be as follows:

$${}^i m_{R_{s',f'}}^{\Omega}[F_{s,f}](\overline{P}_i) = \alpha_{(s,f),(s',f')} \cdot \text{bel}_{IA}^{\Omega}[F_{s',f'}](P_i); \quad (18)$$

$${}^i m_{R_{s',f'}}^{\Omega}[F_{s,f}](\Omega) = 1 - {}^i m_{R_{s',f'}}^{\Omega}[F_{s,f}](\overline{P}_i). \quad (19)$$

With  $\text{bel}_{IA}^{\Omega}[F_{s',f'}](P_i)$  be the credibility measure about the assumption that the person  $P_i$  is the real identity of the face  $F_{s',f'}$ . It is derived from the mass function 17 using the equation 3.  $\overline{P}_i$  represents the union of all elements of  $\Omega$  except  $P_i$ . This mass function associates belief degree to two sets; the incertitude  $\Omega$  and to opposite of a person  $\overline{P}_i$ . It is called contesting strategy because it could reduce the pertinence degree of the metadata  $P_i$  and can't increase it. The coefficient  $\alpha_{(s,f),(s',f')}$  is computed as follows:

$$\alpha_{(s,f),(s',f')} = \Phi_d(d(I_s, I_{s'})) \cdot \Phi_t(t(I_s, I_{s'})) \cdot (1 - \Phi_h(h(R_{s,f}, R_{s',f'}))). \quad (20)$$

With:



- $d(I_s, I_{s'})$  be the geographical distance between  $I_s$  and  $I_{s'}$ .
- $t(I_s, I_{s'})$  be the time between time-taken of  $I_s$  and time-taken of  $I_{s'}$ .
- $h(R_{s,f}, R_{s',f'})$  be the similarity degree between  $R_{s,f}$  and  $R_{s',f'}$ .
- $\Phi_d$  be an increasing function that maps  $d(I_s, I_{s'})$  to interval  $[0, 1]$ . The more the distance is small, the more its value tends towards 1.
- $\Phi_t$  be an increasing function that maps  $t(I_s, I_{s'})$  to interval  $[0, 1]$ . The more the temporal gap is small, the more its value tends towards 1.
- $\Phi_h$  be an increasing function that maps  $h(R_{s,f}, R_{s',f'})$  to interval  $[0, 1]$ . The more similarity measure is high, the more its value tends towards 1.

The functions  $\Phi_d$ , can be a logistic regression functions as follows:

$$\phi_d(\text{dist}(I_s, I_{s'})) = \frac{1}{1 + \exp(-\lambda_d \cdot d(I_s, I_{s'}) + \delta_d)}. \quad (21)$$

The parameters  $\lambda_d$  and  $\delta_d$  permit us to determine a fuzzy threshold between the two opposite situations, near and far geographically. Besides, it is easy to optimize analytically these parameters. We can say the same thing for  $\Phi_t$  and  $\Phi_h$ . These parameters can be updated and learned by a relevance feedback. The mass function  ${}^i m_{R_{s',f'}}^\Omega[F_{s,f}]$  verifies these axioms:

- The more  $t(I_s, I_{s'})$  or  $d(I_s, I_{s'})$  or  $h(R_{s,f}, R_{s',f'})$  is high or  $\text{bel}_{IA}^\Omega[F_{s',f'}](P_i)$  is low, the weak contestation. In other words, the belief degree allocated to uncertainty ( $\Omega$ ) become high, as a result the function  ${}^i m_{R_{s',f'}}^\Omega[F_{s,f}]$  become more neutral (no information).
- The more  $t(I_s, I_{s'})$  and  $d(I_s, I_{s'})$  are low  $h(R_{s,f}, R_{s',f'})$  and  $\text{bel}_{IA}^\Omega[F_{s',f'}](P_i)$  is high, the high contestation.

The masse function 18 is issued from  $R_{s',f'}$  but it is specific to the metadata  $P_i$ . So the absolute mass function issued from  $R_{s',f'}$  could be computed as follows:

$$m_{R_{s',f'}}^\Omega[F_{s,f}] = \bigoplus_{i=1}^N {}^i m_{R_{s',f'}}^\Omega[F_{s,f}]. \quad (22)$$

Finally, the mass function issued from all clothes (to identify  $F_{s,f}$ ) is the fusion of all mass function issued from other faces in other photos:

$$m_{CS}^\Omega[F_{s,f}] = \bigoplus_{(s'=1, \dots, N; f'=1, \dots, z(s')) / (s' \neq s)} m_{R_{s',f'}}^\Omega[F_{s,f}]. \quad (23)$$

**Reinforcing Strategy.** The belief function theory allows us without difficulty to exchange the contesting strategy to reinforcing strategy. The later is used in several works [3, 16, 18] but with other fusion techniques like linear combining and conditional probability. If we want working with reinforcing strategy, the equations 18 and 19 become:

$${}^i m_{R_{s',f'}}^\Omega[F_{s,f}](P_i) = \alpha'_{(s,f),(s',f')} \cdot \text{bel}_{IA}^\Omega[F_{s',f'}](P_i); \quad (24)$$

$${}^i m_{R_{s',f'}}^\Omega[F_{s,f}](\Omega) = 1 - {}^i m_{R_{s',f'}}^\Omega[F_{s,f}](P_i). \quad (25)$$

This mass function associates belief degrees to two sets: the incertitude  $\Omega$  and a person  $\{P_i\}$ . It is called reinforcing strategy because it could increase the pertinence degree of the metadata  $P_i$  and it can't never reduce it. The coefficient  $\alpha'_{(s,f),(s',f')}$  is computed as follows:

$$\alpha'_{(s,f),(s',f')} = \Phi_d(dist(I_s, I_{s'})) \cdot \Phi_t(t(I_s, I_{s'})) \cdot \Phi_h(h(R_{s,f}, R_{s',f'})). \quad (26)$$

In the reinforcing strategy the mass function  ${}^i m_{R_{s',f'}}^\Omega[F_{s,f}]$  verifies these axioms:

- The more  $t(I_s, I_{s'})$  or  $d(I_s, I_{s'})$  or  $h(R_{s,f}, R_{s',f'})$  is high or  $bel_{IA}^\Omega[F_{s',f'}](P_i)$  is low, the weak reinforcing (neutral mass function).
- The more  $t(I_s, I_{s'})$  and  $d(I_s, I_{s'})$  are low and  $h(R_{s,f}, R_{s',f'})$  and  $bel_{IA}^\Omega[F_{s',f'}](P_i)$  are high, the high reinforcing.

Then, we compute the unique mass function issued from clothes recognition  $CS$  as follows (as equations 22 and 23):

$$m_{CS}^\Omega[F_{s,f}] = \bigoplus_{(s'=1, \dots, N; f'=1, \dots, z(s')) / (s' \neq s)} \bigoplus_{i=1}^N {}^i m_{R_{s',f'}}^\Omega[F_{s,f}]. \quad (27)$$

### 3.4 Suggested Metadata List

Finally, For a face  $F_{s,f}$ , we combine the mass functions issued from  $IA$  (equation 17) and from  $CS$  (equation 23 if we use contesting strategy, or equation 27 if we use reinforcing strategy):

$$m_{IACS}^\Omega[F_{s,f}] = m_{IA}^\Omega[F_{s,f}] \oplus m_{CS}^\Omega[F_{s,f}]. \quad (28)$$

Then, we compute the probability pignistic for each element of  $\Omega$  using the equation 11. The suggested metadata list will be ranked according to these probability values. In other words, the metadata  $P_i$  is high ranked than  $P_{i'}$  (with  $i' \in \{1, \dots, N\}$ ) if:

$$BetPm_{m_{IACS}^\Omega[F_{s,f}]}(P_i) \geq BetPm_{m_{IACS}^\Omega[F_{s,f}]}(P_{i'}). \quad (29)$$

This list will be suggested to user for the face  $F_{s,f}$  annotation.

## 4 Experimentation

In this section we present the technique that we have used for clothes detections and segmentation, followed by experiment and evaluation results.

### 4.1 Clothes Detection and Segmentation

This paper focuses on a new method for combining face recognition and clothing similarity. But, we can integrate any available clothing detection and segmentation in our fusion process. We present a simple technique in order to extract the feature vectors of



clothes and computing the similarity between them. If we want to photograph a person, the face is the most interesting and the upper body appears often. That is because we have chosen to utilize the upper body. We use a tool to detect only the frontal faces [17], hence we localize easily the upper body region if a face is detected. For this goal we use a simple technique that is shown in figure 1, it is inspired from [3] and [16]. The face detector provides us the coordinates  $((x_1^f, y_1^f), (x_2^f, y_2^f))$  of two extreme points of face bounding box. Thus, the coordinates of two extreme points of bounding box of upper body region are computed as follows:

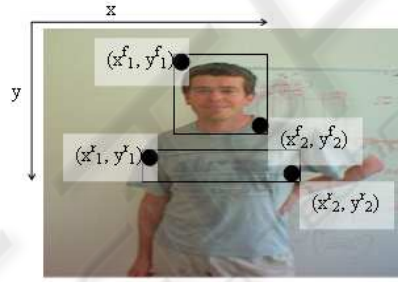
$$x_1^r = \max\{0, x_1^f - \rho_1 \cdot (x_2^f - x_1^f)\}; \quad (30)$$

$$y_1^r = \min\{H, y_2^f + \rho_2 \cdot (y_2^f - y_1^f)\}; \quad (31)$$

$$x_2^r = \min\{W, x_2^f + \rho_1 \cdot (x_2^f - x_1^f)\}; \quad (32)$$

$$y_2^r = \min\{H, y_2^f + \rho_3 \cdot (y_2^f - y_1^f)\}. \quad (33)$$

$W$  and  $H$  denote the width and the height of the photo. The positive values  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  allow us to determinate the position and dimensions of upper body region according to the ones of the detected face ( $\rho_2 < \rho_3$ ).



**Fig. 1.** A detected face and its rectangular part of upper body (a clothes part).

In our experiment, the feature vector of  $R_{s,f}$  is based on color descriptors. The color are quantified into  $HSV$  color space (hue, saturation, value) in 168 bins in such a way that 18 bins for hue, 3 bins four saturation, 3 bins for value and 4 bins for gray level [8]. The percentages of these quantified colors represent the elements of clothes feature vector. To measure the similarity between  $R_{s,f}$  and  $R_{s',f'}$ , we employed this function [5]:

$$h(R_{s,f}, R_{s',f'}) = \frac{\sum_{c=1}^{168} \min(p_{s,f}(c), p_{s',f'}(c))}{\min(\sum_{c=1}^{168} p_{s,f}(c), \sum_{c=1}^{168} p_{s',f'}(c))}. \quad (34)$$

With  $p_{s,f}(c)$  (resp.  $p_{s',f'}(c)$ ) represents the percentage of color  $c$  in  $R_{s,f}$  (resp.  $R_{s',f'}$ ). This similarity function values belong to the interval  $[0, 1]$ . The big value of this function is, the high similarity between clothes.

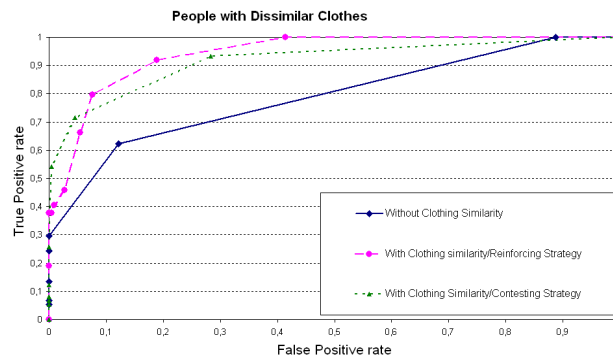
## 4.2 Results

The objective of this evaluation is to validate the choice between reinforcing and contesting strategies. We used an address book that contains 4 people. 90 photos have been collected, each photo showing only one known person. 12 images have been used as model for face recognition tool ( $FR$ ) such that 3 images for each person. In fact we will produce the metadata lists for  $78 = 90 - 12$  faces (all have been detected and, there is no false positive or false negative of face detection tool  $FD$ ). All images are taken in the same day and in the same building (The photographed persons were in dissimilar clothes). The clothes in the photos model of  $FR$  are not taken into consideration. We produced the metadata for the 78 faces. From this initial collection, we extracted another one, where we change the colors of clothes (the extracted rectangular region of upper body). We did that in order to get people with similar clothes (two person with a red clothes and two others with green clothes); in fact, we have used two narrow Gaussian distributions for clothes painting, the center of the first was fixed on the red color and the center of the second was fixed on the green color. Now, we have two collections, the first one with similar clothing and the other being dissimilar. We use each collection separately in order to evaluate these strategies: "without clothing similarity (images analysis  $IA$  only)", "using clothing similarity with reinforcing strategy", and "using clothing recognition with contesting strategy". All photo of our corpus belong to the same event (at the same day and in the same place). So, we believe that for these two collections, the approaches detailed in [18, 3, 16] will give the approximate results to our reinforcing strategy, because they are based only on reinforcing strategy and we have only one event.

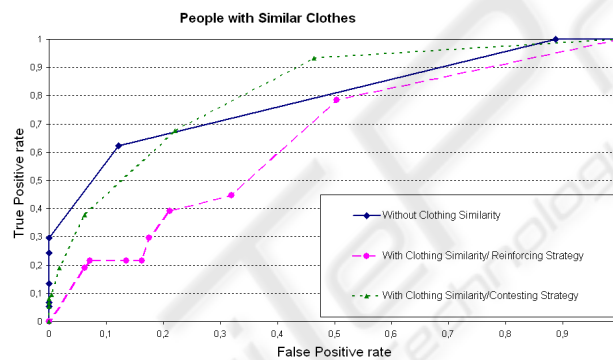
The ROC curves are computed according to metadata scores (pignistic probability). The figure 2 shows the ROC related to the first collection (people in dissimilar clothes). We observe that the use of clothing similarity improves the accuracy of face identification in personal photos and there isn't a big difference between the results of the two strategies reinforcing and contesting. However, the later is better in the beginning (interval  $[0, 0.05]$  of false positive rate). The figure 3 shows the ROC curves related to the second collection (people in similar clothes). We can say that the use of reinforcing strategy is negatively influenced and reduce significantly the system accuracy. But contesting strategy resists more to the noise engendered by similar clothes, further it becomes better from the point "false positive rate 0.22". After all, contesting strategy showed the best results and it is better adapted to photo albums, specially, when we have people with similar appearance.

## 5 Conclusions

In this paper, we have proposed an original approach that uses an external information source for face identification process. On one hand, we attempted to take advantage from the similarity between clothes and the spatiotemporal distances between photos. On the other hand, we tried to use this external information source correctly in order to avoid maximum generation and accumulation of errors. We have seen that, the problem of persons with similar clothes can be reduced without either making clothing similarity unavailable (hard constraint) or less weighted. In addition, the spatiotemporal data was



**Fig. 2.** ROC curves according to thresholds on metadata score, and using the photo collection that contains people in dissimilar clothes.



**Fig. 3.** ROC curves according to thresholds on metadata score, and using the photo collection that contains people in similar clothes.

exploited with a flexible manner and without quantification (no events detection). Note that belief function theory allowed us to deal with our problem efficiently thanks to its capacity of uncertainty modeling and representation. The experimentations showed off clearly the benefit of our approach, the results can be re-ameliorate with the use of an advanced technique of clothes segmentation. In future, we will use this approach for video annotation and present our relevance feedback for parameters updating.

## References

1. M. Abdel-Mottaleb and L. Chen. Content-based photo album management using faces' arrangement. In *IEEE International Conference on Multimedia and Expo, ICME04*, pages 2071–2074, Taipei, Taiwan, June 2004. IEEE.
2. L. Agosto, M. Plu, P. Bellec, and L. Vignollet. Someone : A cooperative system for personalized information exchange. In *International Conference of Enterprise Information Systems*, pages 71–78, Angers, France, 2003.

3. D. Anguelov, K. Lee, S. B. Gokturk, and B. Sumengen. Contextual identity recognition in personal photo albums. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 01–07, June 2007.
4. Y. A. Aslandogan and C. Yu. Multiple evidence combination in image retrieval: Diogenes searches for people on the web. In *Proceedings of the 23rd Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pages 88–95, Athens, Greece, June 2000.
5. A.W.M.Smeulders, M.Worring, S. Santini, A.Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.
6. M. Davis, M. Smith, F. Stentiford, A. Bambidele, J. Canny, N. Good, S. King, and R. Janakiraman. Using context and similarity for face and location identification. In *Proceedings of the IS& T/SPIE 18th Annual Symposium on Electronic Imaging Science and Technology Internet Imaging VII*, page cdrom, San Jose, California, 2006. IS& T/SPIE Press.
7. A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, AMS 38:325–339, 1967.
8. Y. Gong, C. Chuan, and G. Xiaoyi. Image indexing and retrieval using color histograms. *Multimedia Tools and Applications*, 2(2):133–156, March 1996.
9. R. Gossweiler and J. Tyler. PLOG: Easily create digital picture stories through cell phone cameras. In *First International Workshop on Ubiquitous Computing (IWUC 2004)*, pages 94–103, Porto, Portugal, April 2004. INSTICC Press.
10. S. Kharbouche. Fonctions de croyance et indexation multimodale, application à l'identification de personnes dans des albums. Master's thesis, University of Rouen, Rouen, Normandy, France, December 2006.
11. S. Kharbouche and M.Plu. Combination of face detection, face recognition and gender recognition using belief function using evidence theory. In *IEEE International Conference on Information Reuse and Integration*, pages 526–531, Las Vegas, USA, August 2007.
12. M. Naaman, R. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *Proceedings of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005)*, pages 178–187, Denver, Colorado, USA, June 2005.
13. A. Pigeau and M. Gelgon. Incremental statistical geo-temporal structuring of a personal camera phone image collection. In *Proceedings of the 17th International Conference on Pattern Recognition, ICPR'2004*, volume 3, pages 878–881, 2004.
14. G. Shafer. A mathematical theory of evidence. *Princeton University Press*, 1976.
15. P. Smets. Belief induced by the partial knowledge of the probabilities. In D. Heckerman and Al., editors, *Uncertainty in Artificial Intelligence, UAI'94*, pages 523–530, San Mateo, 1994. Morgan Kaufmann.
16. Y. Song and T. Leung. Context-aided human recognition - clustering. In *9th European Conference on Computer Vision ECCV 2006*, pages 382–395, Graz, Austria, May 2006. Springer, LNCS.
17. M. Visani, C. Garcia, and C. Laurent. Comparing robustness of two-dimensional PCA and eigenfaces for face recognition. In *Proceedings of the International Conference on Image Analysis and Recognition*, volume 3212/2004, pages 717–724, 2004.
18. L. Zhang, L.Chen, M. Li, and H. Zhang. Automated annotation of human faces in family albums. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 355–358, New York, NY, USA, 2003. ACM Press.